COMPARATIVE GENOMICS REVEAL ECOPHYSIOLOGICAL ADAPTATIONS OF ORGANOHALIDE-RESPIRING *BACTERIA*

A Thesis
Presented to
The Academic Faculty

by

Darlene Darlington Wagner

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Bioinformatics in the
School of Biology

Georgia Institute of Technology
December, 2012

COMPARATIVE GENOMICS REVEAL ECOPHYSIOLOGICAL

ADAPTATIONS OF ORGANOHALIDE-RESPIRING *BACTERIA*

Approved by:

Dr. Frank E. Löffler, Advisor
Department of Microbiology and Department of Civil and Environmental Engineering,
University of Tennessee

Dr. I. King Jordan
School of Biology
Georgia Institute of Technology

Dr. Konstantinos T. Konstantinidis
School of Civil and Environmental Engineering and School of Biology
Georgia Institute of Technology

Dr. Nicholas Bergman
Genomics Department,
National Biodefense Analysis and Countermeasures Center
Department of Homeland Security

Dr. Thomas DiChristina
School of Biology
Georgia Institute of Technology

Date Approved: November 8, 2012

ACKNOWLEDGMENTS

At the conclusion one of the most rewarding challenges of my life, I reflect upon the collaborative nature of science and how, in the words of Newton, my success rests "upon the shoulders of giants." I am deeply grateful to my thesis advisor, Dr. Frank E. Löffler for seeing the scientist within me and providing many forms of support, often beyond requirements or expectations. Many thanks are due Dr. Konstantinos T. Konstantinidis for his guidance in bioinformatics methods and manuscript writing as well as Dr. King Jordan for his insights into development of bioinformatics methods. Thanks also to my committee members, Dr. Thomas DiChristina and Dr. Nicholas Bergman, as well as other Georgia Tech faculty, Dr. Soojin Yi, Dr. Jung Choi, and Dr. Mark Borodovsky for many helpful discussions. I owe much gratitude also to my remaining fellow student of the Löffler lab, Shandra Justicia-Leon, office-mates, Dr. Rachel Poretsky and Dr. Zohre Kurt; as well as lab members who have moved on: Dr. Benjamin Amos, Dr. Sara Henry Thomas, Dr. Claribel Cruz-Garcia, Dr. Kelly Fletcher, Dr. Kirsti Ritalathi, Dr. Janet Hatt, Dr. Elizabeth Padilla-Crespo, and Melissa Spitzmiller. Additional thanks are due to Kirsti, Janet, and Elizabeth for their helpful discussions and contributions to manuscripts. I am also grateful for the support and kindness of Georgia Tech staff, namely Jenny Eaton, Kevin Roman, and Dean Stephanie Ray. Many thanks also to our collaborators at the University of Toronto, Dr. Laura Hug and Dr. Elizabeth Edwards. Finally, my most heartfelt gratitude goes to my partner, Monica, for her support through my struggles with academics as well as with my new life. Unbounded love and thanks are due to my mom Ruth, my dad Richard, my sisters, Ashley and Allison, my brothers-in-law, Zach and Ben, my uncle Robert, and Aunt Marilyn. I am blessed beyond words to have such family, friends, and colleagues in my life. Many thanks to all!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES – (CONTINUED)

LIST OF FIGURES

Page

LIST OF FIGURES – (CONTINUED)

LIST OF FIGURES – (CONTINUED)

LIST OF SYMBOLS AND ABBREVIATIONS.

Blast	Basic local alignment search tool
BvcA	Vinyl chloride (VC) reductase
CAI	Codon adaptation index
CbrA	Chlorobenzene reductase
CprA	Chlorophenol reductase
cis-DCE	*cis*-1,2-dichloroethene
GC%	Guanine + Cytosine percentage
HMM	Hidden Markov model
HGT	Horizontal gene transfer
ISB	Iron-sulfur binding
MSA	Multiple sequence alignment
NosZ	Nitrous oxide reductase
NOx	Inorganic nitrogen compounds: NO_3^-, NO_2^-, NO, and N_2O
OHRB	Organohalide respiring bacteria (or bacterium)
PCE	Tetrachloroethene
PceA	Tetrachloroethene (PCE) reductase
PCR	Polymerase chain reaction
RDase	Reductive dehalogenase
Tat	Twin arginine translocation pathway
TCE	Trichloroethene
TceA	Trichloroethene (TCE) reductase
TEA	Terminal electron acceptor
VC	Vinyl chloride
VcrA	Vinyl chloride (VC) reductase

SUMMARY

Organohalide-respiring *Bacteria* (OHRB) play key roles in the reductive dehalogenation of natural organohalides and anthropogenic chlorinated contaminants. Reductive dehalogenases (RDases) catalyze the cleavage of carbon-halogen bonds, enabling respiratory energy conservation and growth. Large numbers of RDase genes, a majority lacking experimental characterization of function, are found on the genomes of OHRB. *In silico* genomics tools were employed to identify shared sequence features among RDase genes and proteins, predict RDase functionality, and elucidate RDase evolutionary history. These analyses showed that the RDase superfamily could be divided into proteins exported to the membrane and cytoplasmic proteins, indicating that not all RDases function in respiration. Further, Hidden Markov models (HMMs) and multiple sequence alignments (MSAs) based upon biochemically characterized RDases identified previously uncharacterized members of an RDase superfamily, delineated protein domains and amino acid motifs serving to distinguish RDases from unrelated iron-sulfur proteins. Such conserved and discriminatory features among RDases may facilitate monitoring of organohalide-degrading microbial communities or improve accuracy of genome annotation.

Phylogenetic analyses of RDase superfamily sequences provided evidence of convergent evolution and horizontal gene transfer (HGT) across distinct OHRB genera. Yet, the low frequency of RDase transfer outside the genus level and the absence of RDase transfer between phyla indicate that RDases evolve primarily by vertical evolution or HGT is restricted among related OHRB strains. Polyphyletic evolutionary lineages within the RDase superfamily comprise distantly-related RDases, some exhibiting activities towards the same substrates, suggesting a longstanding history of OHRB adaptation to natural organohalides. Similar functional and phylogenetic

analyses provided evidence that nitrous oxide (N_2O, a potent greenhouse gas) reductase (*nosZ*) genes from versatile OHRB members of the *Anaeromyxobacter* and *Desulfomonile* genera comprised a *nosZ* sub-family evolutionarily distinct from *nosZ* found in non-OHRB denitrifiers. Hence, elucidation of RDase and NosZ sequence diversity may enhance the mitigation of anthropogenic organohalides and greenhouse gases (i.e., N_2O), respectively.

The tetrachloroethene-respiring bacterium *Geobacter lovleyi* strain SZ exhibited genomic features distinguishing it from non-organohalide-respiring members of the *Geobacter* genus, including a conjugative pilus transfer gene cluster, a chromosomal genomic island harboring two RDase genes, and a diminished set of *c*-type cytochrome genes. The *G. lovleyi* strain SZ genome also harbored a 77 kbp plasmid carrying 15 out of the 24 genes involved in biosynthesis of corrinoid, likely related to this strains ability to degrade PCE to *cis*-DCE in the absence of supplied corrinoid (i.e., vitamin B_{12}). Although corrinoids are essential cofactors to RDases, the strictly organohalide-respiring *Dehalococcoides mccartyi* strains are corrinoid auxotrophs and depend upon uptake of extracellular corrinoids via Archaeal and Bacterial salvage pathways. A key corrinoid salvage gene in *D. mccartyi*, *cbiZ*, occurs at duplicated loci adjacent to RDase genes and appears to have been horizontally-acquired from *Archaea*. These comparative genome analyses highlight RDase dependencies upon corrinoids and also suggest mobile genomic elements (e.g., plasmids) are associated with organohalide respiration and corrinoid acquisition among OHRB. In summary, analyses of OHRB genomes promise to enable more complete modeling of metabolic and evolutionary processes associated with the turnover of organohalides in anoxic environments. These efforts also expand knowledge of biomarkers for monitoring OHRB activity in anoxic environments, and will improve our understanding of the fate of chlorinated contaminants.

CHAPTER 1

INTRODUCTION

1.1 Background and Aims of Thesis.

Current knowledge of the organohalide-respiring *Bacteria* (OHRB) is based upon sampling from contaminated environments, pure culture studies, enzymological studies, and gene/genome sequencing. OHRB possess the unique ability to couple anaerobic growth to the reductive cleavage of the carbon-chloride bonds of anthropogenic contaminants of soils, groundwater, and sediments (Holliger et al., 1999; Eekert and Schraa 2001; Smidt and de Vos 2004). OHRB have been characterized through growth in pure cultures on various chlorinated organic compounds as a terminal electron acceptor (TEA). OHRB are distributed across the *Firmicutes* (i.e. low-GC Gram positives), *Chloroflexi*, *δ-Proteobacteria*, and *ε-Proteobacteria*. While some of these organisms were isolated from habitats with no detectable anthropogenic contaminants (Cole et al., 1994; Utkin et al., 1994; Sanford et al., 1996; He et al., 2005), many were derived from sites exhibiting a history of pollution from tetrachloroethene (PCE), trichloroethene (TCE) and other chlorinated solvents, or chlorinated aromatic compounds (Holliger et al., 1993; Nowak et al., 1996; Miller et al., 1997; Rosner et al., 1997; Holliger et al., 1998; Suyama et al., 2001; Wu et al., 2002; He et al., 2003; Sung et al., 2006). Growth studies of OHRB in pure culture have facilitated genome sequencing and enzymological isolation of the reductive dehalogenase (RDase) enzymes mediating the degradation of organohalides. All biochemically-characterized RDases exhibit membrane-associated aspects of functionality (Schumacher and Holliger, 1996; Louie and Mohn, 1999; Pas et al. 2001; Nijenhuis and Zinder, 2005), while all but one of the isolated RDases require a bound

corrinoid (i.e., vitamin B_{12}-related) coenzyme and two iron-sulfur clusters (4Fe-4S) for catalytic activity (Ni et al., 1995; Neumann et al., 1996; Schumacher et al., 1997; Magnuson et al., 1998; Pas et al., 1999). At present, however, a high-resolution crystal structure is lacking for any representative RDase. Functional prediction of uncharacterized RDases by low-resolution structural analyses or molecular modeling is challenging due to the possible membrane-associated nature of RDases and variability in structure of RDase organohalide substrates (e.g., chloroethenes versus chlorobenzenes). As organohalides also occur naturally, they have likely been present in geochemical cycles since the divergence of *Bacteria* and *Archaea* (Gribble, 1998; 2003), suggesting RDases represent a group of evolutionarily ancient enzymes. To elucidate structural, functional, and evolutionary features of RDases, the current studies employ computational comparisons of RDase sequence data.

Given the absence of systems for genetic manipulation among OHRB strains, genome sequencing has provided insights into the physiology, ecology, and evolutionary history of organohalide respiration. Gene and genome sequencing data revealed that experimentally-characterized RDases are encoded as part of a two-gene cluster. RDase A, the corrinoid/iron-sulfur catalytic component is encoded at a 1344-1665 base pair (bp) open reading frame (ORF) while a second, adjacent ORF (225 to 318 bp) encodes an RDase B-subunit predicted to function as a membrane anchor (Hölscher et al., 2004; Smidt and de Vos, 2004). The amino acid sequence of each characterized RDase A-subunit exhibits a conserved set of protein sequence features, an N-terminal twin-arginine translocation (Tat) motif, **RR**x**F**x**K**, and two C-terminal iron-sulfur binding (ISB) motifs, **F**C**x**x**C**x**x**C**x**x**x**C**P** and **C**x_n**C**x_{2-3}**C**x**x**x**C** (Hölscher et al., 2004; Smidt and de Vos, 2004). Gene and protein sequence features among experimentally-characterized RDases have served as a basis for predicting 'putative

RDases' among the rapidly-expanding set of available OHRB genes lacking known function. Moreover, some key groups of OHRB harbor multiple predicted RDase genes, as is the case for strains of *Dehalococcoides mccartyi* which encode up to 106 RDases across five genomes (Kube et al., 2005; Seshadri et al., 2005; McMurdie et al., 2009). A primary goal of bioinformatics is prediction of gene or protein function based upon comparison with a relatively small set of biochemically-characterized genes or proteins (Koonin et al., 1998; Rost, 2002, Friedberg, 2006). Unfortunately, automated genome annotation pipelines do not consistently predict RDase A or B subunit genes on all OHRB genomes, suggesting an enhanced methodology is needed, combining multiple bioinformatics tools (Friedberg, 2006). Thus, meaningful annotation of OHRB genomes and accurate predictions of RDase functionality will depend upon the customization of bioinformatics approaches to RDases and other genes involved in organohalide respiration.

OHRB can be classified according to versatile respiratory metabolism or strict, organohalide-dependent respiration (i.e., obligate OHRB). Genomes of the five *Dehalococcoides mccartyi* strains encode RDases as the only terminal reductase genes, consistent with their description as obligate OHRB (Löffler et al., 2012). OHRB strains with versatile metabolisms, by contrast, may encode oxygen-respiring terminal reductases or various other terminal reductases (e.g., nitrous oxide reductase) as is the case for *Anaeromyxobacter dehalogenans* (Thomas et al., 2008; Thomas et al., 2010). OHRB genomes carry accessory genes critical to RDase functionality, where genes predicted to function in biosynthesis or transport of corrinoids (e.g., vitamin B_{12}) are believed to be essential to organohalide respiration. Obligate OHRB, such as *Dehalococcoides mccartyi*, lack the full set of genes enabling *de novo* corrinoid biosynthesis (Seshadri et al., 2005; Nonaka et al., 2006) while versatile OHRB are

assumed to be capable of synthesizing their own corrinoid (Maphosa et al., 2010). Genome studies have thus revealed a correlation between gene repertoire and respiratory capablities among OHRB. However, the repertoire of functional genes essential OHRB metabolism has not been investigated. The current studies employ and enhance computational methods to identify OHRB genomic features and elucidate their evolutionary history, which may enhance monitoring of organohalide respiratory activity.

Horizontal gene transfer (HGT) has been proposed as a mechanism by which OHRB have adapted to respiration on a broad range of structurally-diverse organohalides. The gene encoding PCE RDase (*pceA*) in *Desulfitobacterium hafniense* strain TCE1 has been shown to be encoded on a self-excising element, suggesting a mobile element for RDase gene transfer (Maillard et al., 2005). Through genome and metagenome sequencing studies of *Dehalococcoides mccartyi*, HGT provides a likely mechanism for the distribution of TCE RDase genes (*tceA*) and VC RDase genes (*bvcA* and *vcrA*) (Krajmalnik-Brown et al., 2007; McMurdie et al., 2011) across geographically-disparate sites. Yet, extensive curation and comparisons serving to detect mobile elements (e.g., transposons, plasmids, etc.) in OHRB genomes are currently lacking. Moreover, evidence is lacking for HGT of accessory functions to RDases, such as genes involved in corrinoid biosynthesis, or HGT crossing different species or genera of OHRB. Customized methods for analysis and curation of protein sequences, genome sequences, and phylogenies may thus provide additional support for the role HGT plays in the evolution of OHRB.

1.2 Hypotheses, Summary of Results, and Discussion.

1.2.1 Research question 1. *Which conserved sequence features of RDase protein sequences enable identification and prediction of functionality?*

Characterized RDases with corrinoid coenzymes likely constitute a new enzyme class with a novel biochemistry (Chapter 2, section 2.2), hence existing bioinformatics tools may not adequately predict RDase functionality among uncharacterized genes (Chapter 2, section 2.3). To this end, a hidden Markov model based upon characterized RDases, validated by conserved RDase protein and gene cluster features, serves to delineate a superfamily comprising the RDases (Hypothesis 1.2.1 A). The RDase HMM retrieved 467 proteins evolutionarily related to RDases. RDase A subunit functionality was predicted in 406 of retrieved proteins, designated class 1, by curation of N-terminal Tat- or Sec-type signal peptides in protein sequences and adjacent gene loci encoding RDase B subunits. The remaining 59 proteins, designated class 2, lacked detectable mechanisms for interaction with the membrane yet could be shown by multiple sequence alignments (MSAs) to belong to the same superfamily as known RDases (Chapter 3, sections 3.3.1 – 3.3.5). Annotation pipelines utilize multiple-sequence-based profiles to annotate genomes of OHRB yet may not delineate the possible multi-domain structure of many RDases (Chapter 2, section 2.3). Hence, the current study sought to show that characterized and predicted RDase proteins exhibit multiple structural or functional domains (Hypothesis 1.2.1 B). Relative to gene annotation in public databases, the HMM and sequence curation approach provided a 5% improvement in prediction of Tat- or Sec- domain-containing RDase A proteins (class 1) and a ~50% improvement in prediction of the class 2 proteins (Chapter 3, section 3.3.6).

1.2.2 Research question 2. *Do genes adjacent to RDases suggest functional or evolutionary specialization?*

The open reading frames (ORFs) adjacent to RDase genes likely encode functions crucial to RDase respiratory function, regulation, horizontal gene transfer (HGT), or interaction with the corrinoid coenzyme (Hypothesis 1.2.2). Genes adjacent to *pceA* and chlorophenol RDases (*cprA*) in *Desulfitobacterium* spp. have been shown in previous studies to be essential to RDase maturation, regulation, or mobilization into transposons (Chapter 2, section 2.3). While RDase B genes have long been proposed to encode membrane anchor subunits to the RDase A catalytic subunit (Chapter 2, sections 2.2 – 2.3), this study demonstrates a conserved membrane topology for RDase B, suggesting critical structural interactions with RDase A (Chapter 3, section 3.3.5). Although the *pceC* and *pceT* regulatory and chaperone genes of *Desulfitobacterium pceA* were not found in all RDase gene clusters, the *pce* cluster of *Geobacter lovleyi* strain SZ harbors *pceC* and *pceT* homologs (Chapter 4, section 4.3.2, Fig. 4.7.1). Unexpectedly, the corrinoid salvage gene, *cbiZ*, was found to be associated with 13 RDase genes on the genomes of *Dehalococcoides mccartyi*, suggesting a role in the functionality of some RDases (Chapter 5, Figs. 5.7.3 and 5.7.5). In summary, all class 1 RDase gene clusters possess RDase B genes while specialized RDase clusters exhibit unique sets of accessory genes (e.g., *pceC*, *pceT*, and *cbiZ*). Together, these RDase accessory genes may provide biomarkers for the monitoring or enhancement of organohalide degradation.

1.2.3 Research question 3. *Do characterized and putative RDases exhibit a polyphyletic evolutionary history?*

Statistical analyses of RDase phylogenies were performed to infer evolutionary relationships among class 1 (membrane-associated) and class 2. Evolutionary scenarios assuming a shared common ancestor for class 1 distinct from class 2 exhibit low probability (< 0.05) (Hypothesis 1.2.3 A). Similarly, characterized RDases cannot be shown to cluster together according to substrate specificity within RDase trees (Hypothesis 1.2.3 B). Lineages representing membrane-associated RDases could not be separated into a single distinct cluster from related class 2 proteins, suggesting class 1 proteins arose multiple times independently (i.e., are polyphyletic). Characterized RDases from distinct evolutionary lineages with overlapping substrate specificities, such as PceA, also exhibited polyphyletic origins (Chapter 3, section 3.3.8). At the same time, RDases on related lineages may exhibit different substrate ranges (Chapter 3, section 3.3.8). Together, these observations suggest RDase functionality or substrate range is not strongly predicted on the basis of clustering within the RDase phylogeny.

1.2.4 Research question 4. *Does the genome of the PCE-respiring Geobacter lovleyi strain SZ possess features distinct from non-organohalide-respiring Geobacter?*

Organohalide respiring and non-organohalide-respiring members within the genus *Geobacter* exhibit differences in gene repertoire and mobile genomic elements (Hypothesis 1.2.4). The *pceA* and *pce* accessory genes on a chromosomal genomic island on *Geobacter lovleyi* strain SZ are lacking in other *Geobacter* spp (Chapter 4, section 4.3.2). The *G. lovleyi* strain SZ genome exhibits other features distinguishing it from other *Geobacter* spp., such as a chromosomal genomic island harboring

conjugative pilus transfer genes and a 77 kbp plasmid (Chapter 4, sections 4.3.3 and 4.3.5). Moreover, *G. lovleyi* strain SZ harbors a diminished set of *c*-type cytochromes and oxidative-stress-responsive genes compared to other *Geobacter* spp. genomes (Chapter 4, section 4.3.4). These findings together suggest that *G. lovleyi* strain SZ has undergone a history of gene losses and acquisitions, possibly related to the acquisition of organohalide respiratory ability, as distinguished from other *Geobacter* spp.

1.2.5 Research question 5. *Do RDases or genes involved in corrinoid acquisition exhibit evidence of horizontal gene transfer (HGT) on OHRB genomes?*

HGT is believed to play a major role in the evolution of OHRB genomes, particularly among *Dehalococcoides mccartyi* (Chapter 2, section 2.3). Scenarios of RDase HGT can be detected by established bioinformatics tools or custom phylogenetic analyses (Hypothesis 1.2.5 A). Customized phylogenetic analyses of RDase sequences indicate at least three instances of RDase HGT crossing distinct genera of OHRB (Chapter 3, section 3.3.9). Analysis of predicted *pceA* genes on the chromosome of *Geobacter lovleyi* strain SZ suggests recent acquisition from outside the *Geobacter* genus (Chapter 4, section 4.3.2). In addition to RDase HGT, genes on OHRB genomes involved in corrinoid biosynthesis or uptake exhibit a history of HGT (Hypothesis 1.2.5 B). Genes encoding the full *de novo* biosynthesis pathway in *Geobacter lovleyi* strain SZ exhibit evidence of lateral transfer from outside the *Geobacter* genus (Chapter 4, section 4.3.7 and chapter 5, section 5.3.2). Two key genes, *cbiB* and *cbiZ*, comprising the incomplete corrinoid biosynthesis pathway in obligate OHRB of the *Chloroflexi* exhibit evidence of lateral transfer from *Archaea* (Chapter 5, section 5.2.3, Fig. 5.7.4). Whereas RDase HGT is likely restricted to closely related strains and

species of OHRB (Chapter 3, section 3.4.8), corrinoid biosynthesis or transport genes on OHRB genomes may have been acquired from taxonomically-distant donors, including *Archaea* (Chapter 5, section 5.4). Yet, OHRB genome features suggest overall evolutionary specialization, evidenced by vertical descent of housekeeping proteins (chapter 5, section 5.3.2) and lack of detectable RDase HGT crossing distinct phyla.

1.2.6 Research question 6. *Do genomes from the genera Anaeromyxobacter and Desulfomonile harbor a nosZ functioning in N_2O respiration?*

Nitrous oxide reductases (NosZ) have been experimentally-characterized in denitrifiers, such as *Bradyrhizobium japonicum* and ammonifiers, such as *Wolinella succinogenes* (Chapter 2, section 2.4). Known *nosZ* genes serve to predict *nosZ* on the genomes of metabolically-versatile OHRB (Hypothesis 1.2.6). NosZ from *Anaeromyxobacter* and *Desulfomonile* possess all residues essential to coordinating the reaction centers found in characterized NosZ (Chapter 6, section 6.3.1). Association of predicted *c*-type and *b*-type cytochromes with *Anaeromyxobacter* and *Desulfomonile nosZ* provide additional evidence of function in a respiratory chain (Chapter 6, section 6.3.3). Hence, the OHRB of *Anaeromyxobacter* and *Desulfomonile* which lack the ability to dentrify, nonetheless may play a key role in N_2O flux.

1.3 References.

Cole, J. R., Cascarelli, A. L., Mohn, W. W., and Tiedje, J. M. (1994). Isolation and characterization of a novel bacterium growing via reductive dehalogenation of 2-chlorophenol. *Applied and Environmental Microbiology* **60**: 3536-3542.

Friedberg, Iddo. (2007) Automated protein function prediction -- the genomic challenge. *Briefings in Bioinformatics* **7**: 225-242.

Gribble, G.W. (1998) Naturally Occurring Organohalogen Compounds. *Accounts of Chemical Research* **31**: 141-152.

Gribble, G.W. (2003) The diversity of naturally produced organohalogens. *Chemosphere* **52**: 289-297.

He, J., Ritalahti, K.M., Yang, K.-L., Koenigsberg, S.S., and Löffler, F.E. (2003). Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* **424**: 62-65.

He, J., Sung, Y., Krajmalnik-Brown, R., Ritalahti, K. M., and Löffler, F.E. (2005). Isolation and characterization of *Dehalococcoides* sp. strain FL2, a trichloroethene (TCE)- and 1,2-dichloroethene-respiring anaerobe. *Environmental Microbiology* **7**: 1442-1450.

Holliger, C., Hahn, D., Harmsen, H., Ludwig, W., Schumacher, W., et al. (1998). *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetrachloroethene and trichloroethene in an anaerobic respiration. *Archives of Microbiology* **169**: 313-321.

Holliger, C., Schraa, G., Stams, A. J. M., and Zehnder, A. J. B. (1993). A highly purified enrichment culture couples the reductive dechlorination of tetrachloroethene to growth. *Applied and Environmental Microbiology* **59**: 2991-2997.

Holliger, C., Wohlfarth, G., and Diekert, G. (1999). Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiology Reviews* **22**: 383-398.

Hölscher, T., Krajmalnik-Brown, R., Ritalahti, K.M., Wintzingerode, F.v., Görisch, H., Löffler, F.E., and Adrian, L. (2004) Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*. *Applied and Environmental Microbiology* **70**: 5290-5297.

Koonin, E. V., Tatusov R. L. and Galperin M. Y. (1998). Beyond complete genomes: from sequence to structure to function. *Current Opinion in Structural Biology* **8**: 355-363.

Krajmalnik-Brown, R., Sung, Y., Ritalahti, K.M., Saunders, F.M., and Löffler, F.E. (2007) Environmental distribution of the trichloroethene reductive dehalogenase gene (*tceA*) suggests lateral gene transfer among *Dehalococcoides*. *FEMS Microbiology Ecology* **59**: 206-214.

Kube, M., Beck, A., Zinder, S.H., Kuhl, H., Reinhardt, R., and Adrian, L. (2005) Genome sequence of the chlorinated compound-respiring bacterium *Dehalococcoides* species strain CBDB1. *Nature Biotechnology* **23**: 1269-1273.

Löffler, F. E., Yan J., Ritalahti K. M., Adrian L., Edwards E. A., Konstantinidis K. T., et al. (2012). *Dehalococcoides mccartyi* gen. nov., sp. nov., obligate organohalide-respiring anaerobic bacteria, relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidetes* classis nov., within the phylum *Chloroflexi*. *International Journal of Systematic and Evolutionary Microbiology* **Apr 27** Epub ahead of print.

Louie, T. M. and Mohn, W. W. (1999). Evidence for a chemiosmotic model of dehalorespiration in *Desulfomonile tiedjei* DCB-1. *Journal of Bacteriology* **181**: 40-46.

Magnuson, J. K., Stern, R.V., Gosset, J.M., Zinder, S.H., and Burris, D.R. (1998). Reductive dechlorination of tetrachloroethene to ethane by a two-component enzyme pathway. *Applied and Environmental Microbiology* **64**: 1270-1275.

Maillard, J., Regeard, C., and Holliger, C. (2005). Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1. *Environmental Microbiology* **7**: 107-117.

Maphosa, F., de Vos, W. M., and Smidt, H. (2010). Exploiting the ecogenomics toolbox for environmental diagnostics of organohalide-respiring bacteria. *Trends in Biotechnology* **28**: 308-316.

McMurdie, P. J., Behrens, S., Müller, J.A., Göke, J., Ritalahti, K.M., Wagner, R. et al. (2009). Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genetics* **5**: e1000714.

McMurdie, P. J., Hug, L., Edwards, E.A., Holmes, S., and Spormann, A.M. (2011). Site-specific mobilization of vinyl chloride respiration islands by a mechanism common in *Dehalococcoides*. *BMC Genomics* **12**: 287.

Miller, E., Wohlfarth, G., and Diekert, G. (1997). Comparative studies on tetrachloroethene reductive dechlorination mediated by *Desulfitobacterium* sp. strain PCE-S. *Archives of Microbiology* **168**: 513-519.

Neumann, A., Wohlfarth, G., and Diekert, G. (1996) Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*. *The Journal of Biological Chemistry* **271**: 16515-16519.

Ni, S., Fredrickson, J.K., and Xun, L. (1995) Purification and characterization of a novel 3-chlorobenzoate-reductive dehalogenase from the cytoplasmic membrane of *Desulfomonile tiedjei* DCB-1. *Journal of Bacteriology* **177**: 5135-5139.

Nijenhuis, I. and Zinder, S. H. (2005). Characterization of hydrogenase and reductive dehalogenase activities of *Dehalococcoides ethenogenes* Strain 195. *Applied and Environmental Microbiology* **71**: 1664-1667.

Nonaka, H., Keresztes, G., Shinoda, Y., Ikenaga, Y., Abe, M., et al. (2006). "Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195. *Journal of Bacteriology* **188**: 2262-2274.

Nowak, J., Kirsch, N. H., Hegemann, W., and Stan, H. J. (1996). Total reductive dechlorination of chlorobenzenes to benzene by a methanogenic mixed culture enriched from Saale river sediment. *Applied Microbiology and Biotechnology* **45**: 700-709.

Raux, E., Schubert, H. L., Roper, J. M., Wilson, K. S., and Warren, M. J. (1999). Vitamin B_{12}: Insights into biosynthesis's mount improbable. *Bioinorganic Chemistry* **27**: 110-118.

Rosner, B. M., McCarty, P. L., and Spormann, A. M. (1997). In vitro studies on reductive vinyl chloride dehalogenation by an anaerobic mixed culture. *Applied and Environmental Microbiology* **63**: 4139-4144.

Rost, B. (2002) Enzyme function less conserved than anticipated. *Journal of Molecular Biology* **318**: 595-608.

Sanford, R. A., Cole, J. R., Löffler, F. E., and Tiedje, J.M. (1996). Characterization of *Desulfitobacterium chlororespirans* sp. nov., which grows by coupling the oxidation of lactate to the reductive dechlorination of 3-chloro-4-hydroxybenzoate. *Applied and Environmental Microbiology* **62**: 3800-3808.

Sanford, R. A., Cole, J.R., and Tiedje, J.M. (2002). Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68**: 893-900.

Schumacher, W. and Holliger, C. (1996). The proton/electron ratio of the menaquinone-dependent electron transport from dihydrogen to tetrachloroethene in "*Dehalobacter restrictus*". *Journal of Bacteriology* **178**: 2328-2333.

Schumacher, W., Holliger, C. Zehnder, A. J. B., and Hagen, W. R. (1997). Redox chemistry of cobalamin and iron-sulfur cofactors in the tetrachloroethene reductase of *Dehalobacter restrictus. FEBS Letters* **409**: 421-425.

Seshadri, R., Adrian, L., Fouts, D. E., Eisen, J. A., Phillippy, A. M., et al. (2005). Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105-108.

Smidt, H., and de Vos, W.M. (2004) Anaerobic microbial dehalogenation. *Annual Review of Microbiology* **58**: 43-73.

Sung, Y., Fletcher, K.E., Ritalahti, K.M., Apkarian, R.P., Ramos-Hernández, N., Sanford, R.A. et al. (2006) *Geobacter lovleyi* sp. nov. strain SZ, a novel metal-reducing and tetrachloroethene-dechlorinating bacterium. *Applied and Environmental Microbiology* **69**: 2775-2782.

Suyama, A., Iwakiri, R., Kai, K,.Tokunaga, T., Sera, N. and Furukawa, K. (2001). Isolation and characterization of *Desulfitobacterium* sp. strain Y51 capable of efficient dehalogenation of tetrachloroethene and polychloroethanes. *Bioscience, Biotechnology, and Biochemistry* **65**: 1474-1481.

Thomas, S. H., Sanford, R. A., Amos, B. K., Leigh, M. B., Cardenas, E., and Löffler, F. E. (2010). Unique ecophysiology among U(VI)-reducing bacteria as revealed by evaluation of oxygen metabolism in *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology* **76**: 176-183.

Thomas, S. H., Wagner, R. D., Arakaki, A. K., Skolnick, J., Kirby, J. R., Shimkets, L. J., Sanford, R. A., and Löffler, F. E. (2008). The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* **3**: e2103.

Utkin, I., Woese, C., and Wiegel, J. (1994). Isolation and characterization of *Desulfitobacterium dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *International Journal of Systematic Bacteriology* **44**: 612-619.

van den Pas, B.A., Jansen, S., Dijkema, C., Schraa, G., Vos, W.M., and Stams, A.J.M. (2001) Energy yield of respiration on chloroaromatic compounds in *Desulfitobacterium dehalogenans*. *Applied and Environmental Microbiology* **67**: 3958-3963.

van den Pas, B.A.v.d., Smidt, H., Hagen, W.R., Oost, J.v.d., Schraa, G., Stams, A.J.M., and de Vos, W.M. (1999) Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in *Desulfitobacterium dehalogenans*. *The Journal of Biological Chemistry* **274**.

van Eekert, M. H. A. and Schraa, G. (2001). The potential of anaerobic bacteria to degrade chlorinated compounds. *Water Science and Technology* **44**: 49-56.

Wu, Q., Milliken, C. E. Meier, J. P., Watts, J. E. M, Sowers, K. R., and May, H. D. (2002). Dechlorination of chlorobenzenes by a culture containing bacterium DF-1, a PCB dechlorinating microorganism. *Environmental Science and Technology* **36**: 3290-3294.

Zhang, Y., Rodionov, D. A., Gelfand, M. S., and Gladyshev, V. N. (2009). Comparative genomic analyses of nickel, cobalt, and vitamin B_{12} utilization. *BMC Genomics* **10**: 78

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of organohalide-respiring *Bacteria* (OHRB).

In a world of elements comprising the building blocks of life, such as carbon or nitrogen, halogens (e.g., Fluorine, Chlorine, etc.) are often overlooked. Over 3,800 naturally-occurring halogen-containing organic compounds (i.e., organohalides) have been discovered (Gribble, 2003), where at least 50% of simple organohalides, such as chloromethanes, can be attributed to natural sources in geochemical cycles (Graedel and Keene, 1996). Natural organohalides may be derived from abiotic sources such as volcanic gases (Jordan et al., 2000) or volatilized chloride from seawater (Graedel and Keene, 1996). Biological sources of organohalides include but are not limited to metabolites from marine organisms (Gribble, 1998; Winterton, 2000; Häggblom et al., 2003) or metabolites from soil microbes (Öberg, 1998; Gribble, 1998). Given that organohalides may have been present on earth since the origins of life (Vetter and Gribble, 2007), microbes have had ample time to evolve enzymatic mechanisms for organohalide detoxification or utilization as a nutrient source. Organohalide respiring *Bacteria* (OHRB) comprise specialized groups of prokaryotic microbes able to couple anaerobic growth to reductive cleavage of carbon-halide bonds in organohalides.

OHRB may also utilize anthropogenic organohalides as a growth substrate and thus are the focus of active research and development for bioremediation methods of organohalide-contaminated sites. The chlorinated aliphatic organohalides, tetrachloroethene (PCE) and trichloroethene (TCE) comprise the primary contaminants of groundwater and soil at up to 40% of U.S. Superfund sites (www.epa.gov/superfund; (Johnson, 1995; Pohl et al., 2008)) while vinyl chloride

(VC) is of great concern due to its acute toxicity and carcinogenicity (http://www.atsdr.cdc.gov/spl/, He, Ritalahti et al., 2003). Chlorinated aromatic compounds, such as hexachlorobenzene, are also a priority for bioremediation efforts due to their toxicity, carcinogenicity, and potential to bioaccumulate in human or animal tissue (www.atsdr.cdc.gov; (Vallack et al., 1998; Vetter and Gribble, 2007)). Given the technical challenges of cleaning up organohalide contaminants in subsurface environments such as groundwater, OHRB are increasingly utilized in bioremediation strategies (Alexander, 1991; Löffler et al., 2006; Tas et al., 2010). The respiratory physiologies of some OHRB strains also show promise in bioelectricity or microbial fuel cell applications (Lovley, 2006; Strycharz et al., 2008; Strycharz et al., 2010). Yet, a more complete understanding of OHRB and their applications may depend upon accurate modeling of their metabolism and evolutionary origins through the tools of genome sequencing and bioinformatics.

The current review highlights what is currently known regarding organohalide respiring *Bacteria* and their impacts on geochemical cycling of compounds of concern. The first section details experimental findings surrounding the biochemistry, ecology, and microbiology of OHRB. The second section describes key genes on OHRB genome sequences likely to play a role in degradation of organohalides and evolutionary adaptations among OHRB. The third section highlights the role played by some OHRB strains in the cycling of inorganic nitrogen. Together, experimental and genomic data for OHRB yield valuable insights into the mitigation of environmental contaminants, modeling of OHRB metabolism or evolution, and understanding the role of microbes in geochemical cycling.

2.2 Ecophysiology of organohalide respiration.

Over the past three decades, isolation and physiological characterization of organohalide respiring *Bacteria* (OHRB) have elucidated the biochemistry of organohalide respiration and served to predict conditions favorable to bioremediation of organohalide-contaminated sites. Pure culture studies of OHRB also enable characterization of organohalide-degrading enzymes and sequencing of genes or genomes. In contrast to respiratory processes which utilize a single substrate of defined structure, e.g. oxygen or nitrate, organohalide respiration encompasses a broad diversity of compounds. Hence, experimental characterization of organohalide respiring organisms or enzyme systems can often yield surprising or unexpected insights.

Redox conditions and prevailing microbial processes in subsurface environments are conducive to organohalide degradation involving the reductive cleavage of carbon-halogen bonds. Cobalamin (vitamin B_{12}), coenzyme F_{430}, and other metalloporphyrins have been shown to catalyze the reductive dechlorination of PCE, TCE, and chloromethanes in subsurface environments dominated by methanogens or acetogens (Gantzer and Wackett, 1991). The presence of cobalamin or analogous corrinoid compounds among many groups of anaerobic *Bacteria* (Gantzer and Wackett, 1991; Fetzner and Lingens, 1994) and the abundance of electron donors, such as hydrogen (H_2) under fermentative or methanogenic conditions are favorable to organohalide degradation. Many *in situ* organohalide degradation processes do not provide energy for microbial growth and have been termed "co-metabolic reductive dehalogenation" (Fantroussi et al., 1998). Cometabolic organohalide dehalogenation is often slow, seldom faster than 200 nmol Cl^- (mg protein)$^{-1}$ per day for chlorinated compounds (Holliger and Schumacher, 1994; van

Eekert and Schraa, 2001). By contrast, organohalide-degradation processes characteristic of OHRB couple the reductive cleavage of the carbon-halogen bond to the production of ATP (Figure 1 A) and exhibit faster reaction rates. For instance, during experimental characterization of source cultures for some of the first OHRB isolates, *Dehalococcoides mccartyi* strain 195 and *Sulfurospirillum multivorans*, respiratory dehalogenation was shown to proceed at least 2 to 3 orders of magnitude faster than cometabolic processes (Tandol, DiStefano et al., 1994; Scholz-Muramatzu et al., 1995; van Eekert and Schraa, 2001). Hence, initial culturing and isolation of OHRB strains yielded valuable insights into organohalide degradation as an energy-yielding, respiratory process and its possible applicability to bioremediation.

Reductive dehalogenase (RDase) enzymes are the evolutionarily-specialized enzyme systems isolated from OHRB shown to couple the cleavage of the carbon-halogen bond to respiratory growth. Coupling of carbon-halogen bond reduction to the production of ATP depends upon a key set of RDase functional components, iron-sulfur clusters, corrinoids, and a means for association with the cytoplasmic membrane (Figures 1 A and 1 B). Theoretical studies and characterization of cell- or enzyme-free extracts suggest that the cobalt ion in corrinoid compounds binds the carbon atom from which the halide substituent is cleaved (Banerjee and Ragsdale, 2003; Kliegman and McNeill, 2008). The two iron-sulfur centers in RDases are believed to each donate one electron to the cobalt ion of the cobalamin coenzyme as part of the redox cycle from Co(III), to Co(I) (Neumann et al., 1996). The PCE RDase (PceA) from *Dehalobacter restrictus* exhibited midpoint redox potentials E_m = -480 mV in the two 4Fe-4S centers and E_m = -350 mV in the cobalamin coenzyme (Schumacher, Holliger, et al., 1997). Similar chemical properties have been elucidated for the PceA from *Sulfurospirillum multivorans* (Neumann et al., 1996), the chlorophenol RDase (CprA)

from Desulfitobacterium dehalogenans (van den Pas et al., 1999), and the TCE RDase (TceA) from *Dehalococcoides mccartyi* strain 195 (Magnuson et al., 1998). The only RDase shown to lack corrinoid and iron-sulfur clusters was purified from the *δ-Proteobacterium*, *Desulfomonile tiedjei*, which appears to employ a heme cofactor for enzymatic activity towards 3-chlorobenzoate (Ni et al., 1995). Although a crystal structure is lacking for any RDase, the occurrence of corrinoid and iron-sulfur redox centers across a majority of RDases suggests a conserved pathway for electron transfer to organohalide substrates.

RDase enzyme interaction with the cytoplasmic membrane is the most crucial feature of organohalide degradation as an energy-yielding process. The *Dehalobacter restrictus* PceA translocates one proton across the cytoplasmic membrane for each chloride released in the presence of H_2 as an electron donor (Schumacher and Holliger, 1996). The *Desulfitobacterium dehalogenans* CprA was similarly shown to be membrane-associated but the orientation of the active site, inside versus outside the cytoplasm, could not be determined (van den Pas et al., 2001). Freeze fracture studies have shown that the PceA of *Sulfurospirillum multivorans* (Figure 1 B) is oriented on the periplasmic side of the cytoplasmic membrane (John et al., 2006). The TceA of *Dehalococcoides mccartyi* strain 195 was also inferred to have a periplasmic orientation (Nijenhuis and Zinder, 2005). The heme-binding 3-chlorobenzoate RDase of *Desulfomonile tiedjei* is also believed to generate a proton gradient across the membrane (Louie and Mohn, 1999). Hence, biochemically-characterized RDases from representatives of four phyla, *ε-Proteobacteria*, *δ-Proteobacteria*, *Firmicutes*, and *Chloroflexi*, all exhibit membrane-associated coupling of organohalide degradation to cellular growth.

Despite the energy available in the carbon-halogen bonds of substrates, organohalide respiration is an unexpectedly inefficient process. Differences between the midpoint redox potentials for the RDase corrinoid center (E_m < -350 mV) and the carbon-chloride bond (+360 mV < E_m < +580 mV) (Vogel et al., 1987) would suggest organohalide respiration is equal to nitrate reduction in terms of energetic favorability (Gantzer and Wackett, 1991). With H_2 as a reductant, the dechlorination of PCE to *cis*-1,2-dichloroethene (*cis*-DCE) yields a free energy change (ΔG') of -189 kJ/mol and the removal of the chloride substituent from the aromatic, 3-chlorobenzoate, yields ΔG' of -137 kJ/mol (Holliger et al., 1999). This implies a theoretical energy yield of at most 2 ATP per chloride removed from various chlorophenols or chlorobenzoates (Fantroussi et al., 1998; van den Pas et al., 2001) and 2.5 ATP per chloride removed from PCE (Fantroussi et al., 1998; Holliger et al., 1999). However, the relatively low growth yields of OHRB strains suggests that no more than 1 ATP, or as little as 1/3 ATP (van de Pas et al., 2001), is generated for the cleavage of each carbon-halogen bond. It has been proposed that a reverse flow of two electrons is required to return the oxidized Co(III) in the cobalamin cofactor to the Co(I) oxidation state, thus diminishing energy yield (John et al., 2006). Yet, despite inefficiencies in organohalide respiration, OHRB utilize a broad range of structurally-diverse organohalide substrates across widely-dispersed anaerobic environments.

OHRB isolates have been key to elucidating the environmental distribution and substrate range of organohalide respiration where over 30 OHRB, representing low-GC Gram positives (*Firmicutes*), *Chloroflexi*, *δ-Proteobacteria*, and *ε-Proteobacteria* have been described in pure culture (Figure 2). Many OHRB exhibit an affinity for halogenated aromatic compounds (e.g., chlorobenzenes) as growth substrates. Among the first isolated OHRB were from the *δ-Proteobacteria*, where *Desulfomonile tiedjei*

strain DCB-1 from methanogenic sewage sludge (Shelton and Tiedjei, 1984; Mohn and Kennedy, 1992) and *Anaeromyxobacter dehalogenans* strain 2CP-1 from stream sediment (Cole et al., 1994), respire on 3-chlorobenzoate and 2-chlorophenol, respectively. Other OHRB isolates respiring chlorophenols or other chloroaromatics represented *Firmicutes* such as *Desulfitobacterium hafniense* strain DCB-2 from municipal sludge (Madsen and Licht, 1992), *D. dehalogenans* strain JW/IU-DC1 from pristine freshwater sediments (Utkin et al., 1994), and *D. chlororespirans* strain Co23, from compost (Sanford et al., 1996). Compared to the *δ-Proteobacteria* or *Firmicutes*, *Chloroflexi* OHRB utilize a broader and more structurally-diverse range of chloroaromatics. For example, *Dehalococcoides mccartyi* strain CBDB1, derived from contaminated river sediment (Nowak et al., 1996), respires on chlorobenzenes (Adrian et al., 2007), highly-chlorinated congeners of dibenzo-*p*-dioxins (Bunge et al., 2003) and biphenyls (Adrian et al., 2009).

Other OHRB, distributed across all four phyla shown in Figure 2, have been characterized for their specificity towards chlorinated aliphatics. The *ε-Proteobacterium*, *Sulfurospirillum multivorans*, was isolated from an activated sludge enrichment (Scholz-Muramatsu et al., 1995) and grows by the reduction of PCE to *cis*-DCE. *Dehalobacter restrictus* PER-K23, from the *Firmicutes*, was isolated from contaminated river sediment (Holliger, Schraa et al., 1993) and grows exclusively by the reduction of PCE to *cis*-DCE (Holliger, Hahn, et al., 1998). Other PCE to *cis*-DCE degrading *Firmicutes*, *Desulfitobacterium* sp. PCE-S (Miller et al., 1998) and *Desulfitobacterium hafniense* Y51 (Suyama et al., 2002), were isolated from PCE-contaminated soils (Miller et al., 1997; Suyama et al., 2001). Two members of the *δ-Proteobacteria* also grow by reduction of PCE to *cis*-DCE, *Desulfuromonas*

michiganensis, with strains representing both pristine and chloroethene-contaminated aquifers (Sung et al., 2003) and *Geobacter lovleyi* strain SZ from stream sediments (Sung et al., 2006). As for chlorinated aromatics, OHRB of the *Chloroflexi* utilize the widest diversity of chlorinated aliphatic compounds where the ability to degrade PCE and TCE completely to the non-toxic end product, ethene (Figure 3), is found only among the *Chloroflexi*. *Dehalococcoides mccartyi* strain 195, from a methanogenic culture (DiStefano et al., 1991; Tandol et al. 1994), degrades PCE and TCE to VC and 1,2-dichloroethane to ethene (Maymó-Gatell et al., 1997; Maymó-Gatell et al., 1999) (Figure 3). *Dehalococcoides mccartyi* strains BAV1 (He et al., 2003) and VS (Rossner et al., 1997; Müller et al., 2004), both from contaminated aquifers, efficiently complete the degradation of *cis*-DCE or VC to ethene. Thus, differences in chlorinated substrates utilized by different OHRB strains reflect variations in RDase enzyme substrate ranges. A critical area of research into OHRB metabolism is the prediction of organohalide substrate utilization based upon DNA or protein sequences. In addition, it is not clear how differences in taxonomic affiliation or environment influence OHRB metabolism or evolution.

2.3 Genomics of organohalide respiration.

OHRB genome sequencing promises crucial insights into enzyme systems, metabolic pathways, and evolutionary history of organohalide respiration. Most OHRB, particularly *Dehalococcoides mccartyi* strains, are difficult to grow in pure culture and lack systems for genetic manipulation (Maphosa et al., 2010; Löffler et al., 2012), When wet-lab characterization is not feasible, genome sequencing and the computational sequence comparison tools of bioinformatics together address knowledge gaps in OHRB metabolism, physiology, and evolution. The primary

methodology of bioinformatics genome analysis is the prediction of gene functions through computational database search and pairwise sequence alignment tools such as Blast (Altschul et al., 1990; Altschul et al., 1997; Friedberg, 2006). Yet, description of distinctive sequence features within a family or superfamily of proteins requires bioinformatics methods based upon multiple sequence alignments (MSAs) and hidden Markov models (HMMs), which outperform pairwise methods such as Blast (Koonin et al., 1998; Park et al., 1998; Durbin et al., 1998; Todd et al., 2001; Edgar, 2004). As MSAs serve to reconstruct the shared evolutionary history of groups of similar sequences (Park et al., 1998), MSAs are indispensable to prediction of protein functional domains (Finn et al., 2008) or phylogeny reconstruction using tools such as Phylip (Felsenstien, 1989). For high-throughput prediction of gene function, genome databases such as RefSeq incorporate protein function prediction profiles such as PFAM into their genome annotation pipelines (Pruitt et al., 2007; Finn et al., 2008). Yet, inadequate curation of multi-domain proteins and the occasional lack of functional homology between otherwise highly-similar sequences may generate erroneous gene function predictions (Reeck et al., 1987; Friedberg, 2006). Despite current limitations, data obtained by comparative genome analyses enables further experimental characterization of OHRB strains through molecular tools such as polymerase chain reaction (PCR) (Smidt, et al., 2000; Maphosa et al., 2010). In this manner, increasing availability of OHRB genomes may serve to predict functionality of organohalide respiratory genes, reconstruct OHRB evolution, or assay OHRB activity in the environment.

Sequence comparisons among gene loci for biochemically-characterized RDases reveals a conserved set of features associated with the functional components of organohalide respiration. DNA sequences from RDase experimental studies

revealed a 1344-1665 base pair (bp) gene encoding the RDase catalytic subunit (RDase-A) and an adjacent smaller gene (225 to 318 bp) encoding a membrane-bound protein (B-subunit) believed to function as a membrane anchor to the A-subunit (van den Pas et al., 1999; Suyama et al., 2002; Maillard et al., 2003) (Figure 4 A). For example, the PCE RDase protein (PceA) of *Dehalobacter restrictus* was found to be encoded as part of a gene cluster, *pceA-pceB*, with the gene encoding the predicted membrane anchor protein, PceB (Maillard et al., 2003). MSAs of amino acid sequences revealed an N-terminal consensus sequence, **RR**x**F**x**K** (Figure 4 A), in characterized RDase A proteins characteristic of the twin-arginine translocation (Tat) mechanism for protein export outside the cytoplasmic membrane (van de Pas et al., 1999; Suyama et al., 2002; Maillard et al., 2003; Berks et al., 2005; Krajmalnik-Brown et al., 2004). The two 4Fe-4S (or 3Fe-4S) iron-sulfur centers in isolated RDases correspond to two highly-conserved C-terminal iron-sulfur-binding motifs (Figure 4 A, denoted ISBs) readily detectable in RDase A protein MSAs (van de Pas et al., 1999; Suyama et al., 2002; Maillard et al., 2003; Smidt and de Vos, 2004). The first ISB site of RDase A proteins has the amino acid consensus, **FC**xx**C**xx**C**xxx**CP**, typical of bacterial ferredoxins while the second ISB motif has consensus atypical among iron-sulfur proteins, **C**x_n**C**x_{2-3}**C**xxx**C** (Smidt and de Vos, 2004; Hölsher et al., 2004). Although biochemically-characterized RDases have been shown to have a bound corrinoid coenzyme (Banerjee et al., 2003; Krastokina et al., 2001; Kraütler et al., 2003; Maillard et al., 2003; Siebert et al., 2002), known corrinoid-binding motifs, such as **D**x**H**xx**G**x_n**S**x**L**x_n**GG**, are lacking in the amino acid sequences of all characterized RDases (Ludwig and Matthews, 1997; van den Pas et al., 1999; Smidt and de Vos, 2004; Hölsher et al., 2004). RDase B protein sequences possess two to three predicted transmembrane helices, and a conserved motif, **WY**x**W** (van de Pas et

al., 1999; Suyama et al., 2002; Maillard et al., 2003; Hölscher et al., 2004; Krajmalnik-Brown et al., 2004). Yet, neither the orientation of B-proteins within the membrane nor amino acid residues mediating interaction between the RDase A and RDase B has been investigated. In summary, RDase functional components enabling export outside the cytoplasm (i.e., Tat motif), interaction with the membrane (RDase B subunit), and storage/transfer of electrons (iron-sulfur centers), are associated with sequence features. Unfortunately, given the paucity of RDase structural data, more detailed knowledge of RDase sequence features, such as residues involved in coenzyme binding or catalysis, are not available.

Whole-genome sequencing provides a basis for predicting the occurrence and functionality of RDase genes and their relationship to organohalide substrate utilization by OHRB. Sequences from biochemically-characterized RDase gene clusters (*pceA-pceB*, *cprB-cprA* (Figure 4 A)) serve as a basis for predicting genomic RDase genes which lack experimental confirmation of function (i.e., putative RDases). As a result, over 200 genes encoding putative RDase A (*rdhA*) with adjacent RDase B (*rdhB*) have been identified on completed OHRB genomes, where the gene symbol "*rdh*" denotes "reductive dehalogenase homologous." Considering that it is uncertain whether all RDases share evolutionary and functional homology, the term "RDase homologous" may actually mean "RDase similar" (Reeck et al., 1987). Despite inconsistencies in terminology describing RDase sequence features, computational analyses have revealed trends in genome counts of RDases relevant to OHRB physiology. For example, the iron-respiring *Desulfitobacterium hafniense* strain DCB-2 harbors one characterized *cprB-cprA* (chlorophenol RDase), four intact *rdhA-rdhB* loci, and two additional predicted RDases (Kim et al., 2012). The chlorophenol-respiring microaerophile, *Anaeromyxobacter dehalogenans* strain 2CP-C, harbors two

putative RDases, albeit lacking RefSeq annotation suggestive of dehalogenase function (Sanford et al., 2002; Thomas et al., 2008; Thomas et al., 2010). Genomes of the obligate organohalide respirers of *Dehalococcoides mccartyi* encode larger numbers of characterized and putative RDases, where RDase gene counts are generally consistent among expert curators. In total, *D. mccartyi* strain 195 harbors 17 RDases (Seshadri et al., 2005), strain CBDB1, 32 (Kube et al., 2005), strain BAV1, 11 (McMurdie et al., 2009), strain VS, 36 (McMurdie et al., 2009), and strain GT, 20 (Sung et al., 2006). The increased numbers of RDase genes on *Dehalococcoides* appear to be related to their expanded range of organohalide substrates in comparison to other OHRB groups (Maphosa et al., 2010; Löffler et al., 2012). As additional OHRB genomes become available, more rigorous methodology for computational prediction of RDases must be developed. Ideally, shared amino acid motifs (Bork and Koonin, 1996; Koonin et al., 1998), protein sequence identities > 30% (Park et al., 1998; Friedberg, 2006), and conserved gene neighborhoods (Dandekar et al., 1998; Friedberg, 2006), would together serve as criteria for consistent prediction of RDases. In summary, automated gene annotation and expert gene curation have uncovered useful information regarding the differences between versatile and obligate OHRB strains; however, the development of better computational tools for inferring RDases may enable further discoveries, including prediction of RDase substrate range.

OHRB whole genome sequencing has revealed that RDase A and RDase B genes often comprise clusters with additional genes playing possible accessory roles to RDase functionality or RDase horizontal gene transfer (HGT). For example, *Desulfitobacterium hafniense pceA* belongs to a four-gene cluster, *pceA-pceB-pceC-pceT* (Figure 4 B), where *pceC* is believed to encode a regulatory protein (Futagami et al., 2006) and *pceT* has been found essential to PceA protein maturation (Morita et al.,

2009). Similarly, the chlorophenol RDase gene (*cprA*) from *Desulfitobacterium dehalogenans* belongs to a cluster, *cprT-cprK-cprZ-cprE-cprB-cprA-cprC-cprD* (Figure 4 C) where *cprK* regulates expression of CprA (Pop et al., 2004), and *cprT*, *cprE*, and *cprD* may be involved in CprA maturation (Smidt, Leest, et al., 2000). RDase genes from *D. mccartyi* strain 195 (Seshadri et al., 2005) and strain CBDB1 (Kube et al., 2005) are often localized in clusters with uncharacterized genes, *rdhC*, *rdhD*, and/or *rdhR*, believed function in regulation (Figure 4 D). In addition, the TCE RDase gene, *tceA*, and three putative RDase genes on the *D. mccartyi* strain 195 genome possess adjacent transposases and tRNA-gene integration sites (Seshadri et al., 2005) suggestive of mechanisms for gene mobility. Genomic regions harboring the VC RDases, *vcrA* from *D. mccartyi* strain VS and *bvcA* from *D. mccartyi* strain BAV1, have adjacent tRNA/tmRNA direct repeats and exhibit guanine-cytosine (GC) content uncharacteristic of their host genomes (McMurdie et al., 2009; McMurdie et al., 2011). Moreover, sequencing of *tceA-tceB* and *vcrA-vcrB* from mixed culture PCR amplicons suggests these genes are widely distributed among *Dehalococcoides* populations (Krajmalnik-Brown et al., 2007; McMurdie et al., 2011). That RDase genes with required accessory genes are harbored on mobile elements (e.g., transposases) suggests a mechanism for rapid adaptation of OHRB to organohalide substrates of diverse structures.

As cleavage of the carbon-halogen bond requires the input of electrons, OHRB genomes encode respiratory enzyme complexes making reducing equivalents available to RDases from H_2 or various organic carbon compounds. For example, *Desulfitobacterium hafniense* strain Y51 uses pyruvate and lactate as electron donors for organohalide respiration (Suyama et al., 2001) and accordingly harbors multiple pyruvate:ferredoxin oxidoreductase genes (Nonaka et al., 2006). *Dehalococcoides*

mccaryti, by contrast, are dependent of H_2 as a source of reducing equivalents during organohalide respiration and each harbor five non-identical hydrogenase gene clusters, *hup*, *ech*, *vhu*, *hyc*, and *hym* (Seshadri et al., 2005; Kube et al., 2005; Morris et al., 2006; Morris et al., 2007; Löffler et al., 2012). Expression studies show *hup*, encoding periplasmic nickel-iron hydrogenase, is the most strongly upregulated of *Dehalococcoides* hydrogenases during respiration on PCE (Rahm et al., 2006; Morris et al., 2007) suggesting function in coupling H_2 oxidation to organohalide reduction. Other OHRB, such as *Geobacter lovleyi* strain SZ, use acetate as a source of reducing equivalents during organohalide respiration (Sung et al., 2006), but the enzyme system(s) coupling acetate oxidation to organohalide reduction are currently unknown. Yet, availability of OHRB genomes enables further computational analyses for genes conserved across OHRB as well as the development of molecular tools (e.g., reverse transcriptase PCR) as a means of identifying genes functionally-linked to RDases.

Given that the corrinoid coenzyme is essential to RDase function, genes mediating pathways for corrinoid biosynthesis or corrinoid uptake are likely essential to OHRB. Up to 30 enzymatic functions are required for *de novo* biosynthesis of cobalamin or related corrinoids (Roessner et al., 2002) where genes enabling the full pathway are found on the genomes of *Desulfitobacterium hafniense* strains and *Geobacter lovleyi* strain SZ (Nonaka et al., 2007; Kim et al., 2012; Wagner et al., 2012). Accordingly, *Geobacter lovleyi* strain SZ respires PCE in pure cultures without amended vitamin B_{12} or in co-cultures with corrinoid auxotrophs (Wagner et al., 2012; Yan et al., 2012). By contrast, *Dehalococcoides mccartyi* strains require amendments of corrinoids in the form of commercial vitamin B_{12} or co-culturing with organisms capable of *de novo* corrinoid biosynthesis (He et al., 2007; Johnson et al., 2009; Yan et al., 2012; Men et al., 2012; Löffler et al., 2012). While *D. mccartyi* genomes carry a

reduced set of corrinoid biosynthesis genes, only nine of the nearly 30 genes required for *de novo* biosynthesis (Seshadri et al., 2005; Nonaka et al., 2006), they exhibit regulatory control dependent upon corrinoid source, i.e., commercial vitamin B_{12} vs. co-culture (Johnson et al., 2009; Men et al., 2012). Given that many OHRB are associated with microbial communities dominated by corrinoid-rich methanogens or acetogens, OHRB may have evolved novel mechanisms for bypassing *de novo* corrinoid biosynthesis and utilizing extracellular sources.

2.4 OHRB in cycling of NOx compounds.

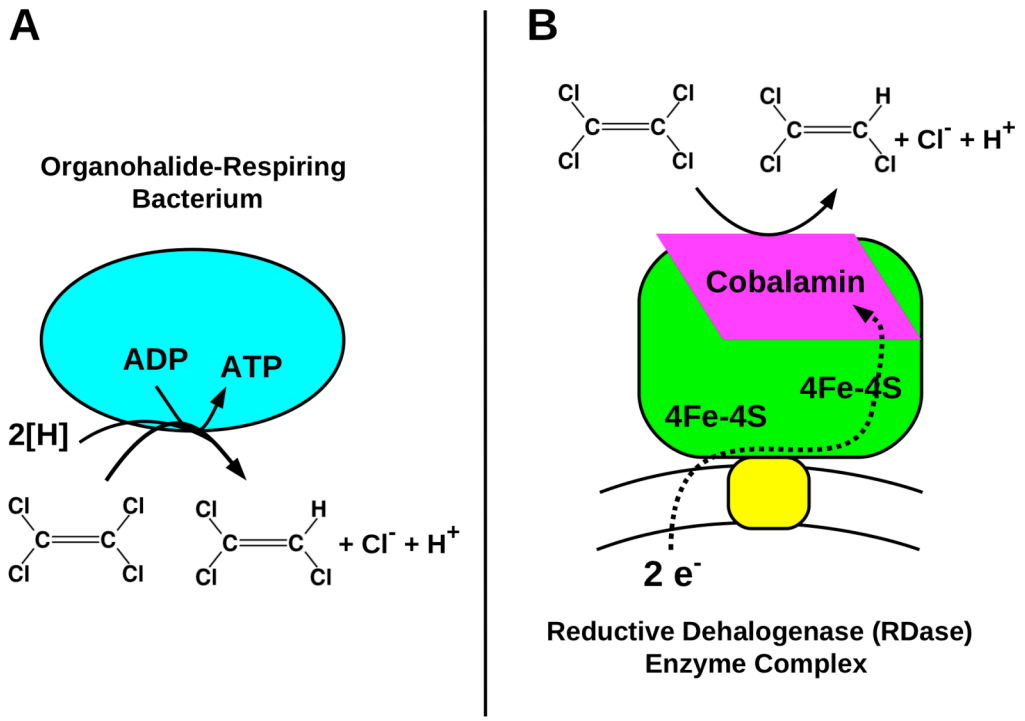
Microbial activities in response to anthropogenic disturbances to the global nitrogen cycle have serious implications to greenhouse gas emissions as well as the availability of drinking water. Industrial production of ammonium (NH_4^+) by the Haber-Bosch process and various intensive agricultural practices are among the leading human activities leading to imbalances in the natural geochemical cycling of nitrogen (Galloway, 2004). Nitrate (NO_3^-) may comprise the sole groundwater contaminant in agricultural areas or may be present in mixtures with organohalides and pesticides (Nolan and Ruddy, 1996; Squillace et al., 2002), both situations posing significant health and developmental risks to young children (Manassaram et al., 2006). Extensive applications of inorganic nitrogen (NH_4^+ and NO_3^-) in agriculture stimulate various microbial anaerobic respiratory processes: ammonification ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$), nitrification ($NH_4^+ \rightarrow NO_2^-$), complete denitrification ($NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$), and partial denitrification ($NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O$) (Zumft, 1997; Philippot, 2002). Each of these processes, particularly partial dentrification, produce reactive intermediates, nitric oxide (NO) and nitrous oxide (N_2O). N_2O acts as a greenhouse gas over 300 times more potent than CO_2 (Wang et al., 1976; Lashof and

Ahuja, 1990; Finlayson-Pitts and Pitts, 2000; Stein and Yung, 2003; McLain and Martens, 2005) while both NO and N_2O contribute to destruction of stratospheric ozone (Cicerone et al., 1987; Crutzen et al., 1970; Ravishankara et al., 2009). The enzyme, nitrous oxide reductase (NosZ), catalyzes the final step of denitrification, reducing reactive N_2O to inert N_2 (Zumft, 1997; Zumft and Körner, 2007; Zumft and Kroneck, 2007). Although experimental characterization of NosZ has been primarily limited to complete denitrifiers of the *α*-, *β*-, and *γ-Proteobacteria* (Phillippot, 2002; Zumft and Körner, 2007; Zumft and Kroneck, 2007), a functional NosZ enzyme has also been characterized from the ammonifying ε-Proteobacterium *Wolinella succinogenes* (Payne et al., 1982; Simon et al., 2004) and the Gram-positive dentrifier, *Geobacillus thermodenitrificans* (Liu et al., 2008). Hence, processes essential to the health and sustainability of the global nitrogen cycle appear to be widespread among subsurface microbes.

Experimental evidence and genome comparisons suggest metabolically-versatile OHRB may play key roles in converting various inorganic contaminants of the subsurface and mitigating emissions of the greenhouse gas, N_2O. Metabolically-versatile OHRB of the *Anaeromyxobacter* as well as *Geobacter lovleyi* strain SZ, were characterized upon isolation for their ability to perform ammonification ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$). Consistent with these observations, genomes of *Geobacter lovleyi* strain SZ and *Anaeromyxobacter dehalogenans* strains CP-C, 2CP-1, and K harbor genes encoding proteins similar to periplasmic nitrate reductases (NapA) (Marietou et al., 2005; Kern and Simon, 2009) and ammonia-forming cytochrome *c* nitrite reductases (NrfA) (Kern and Simon, 2009; Thomas et al., 2008). Comparative genomics analysis of *A. dehalogenans* strains 2CP-C, 2CP-1, and K also revealed a putative nitrous oxide reductase gene (*nosZ*), suggesting *Anaeromyxobacter* derives energy for growth from

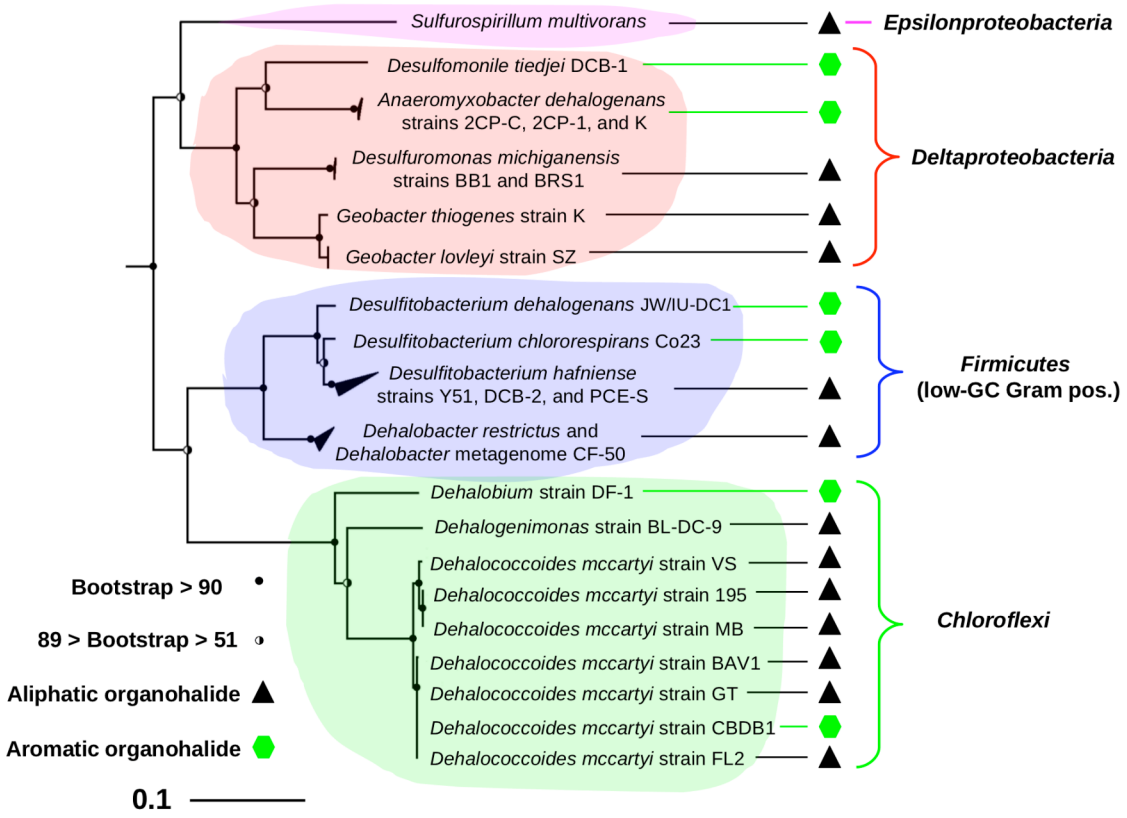
the reduction of N_2O to N_2 (Thomas et al., 2008). Although nitric oxide forming nitrite reductases (NirK or NirS) are absent from *Anaeromyxobacter* genomes, three non-identical copies of putative nitric oxide reductases (*nor*) are present across *Anaeromyxobacter dehalogenans* genomes (Thomas et al., 2008). Hence, detoxification of highly-reactive NO, as well as respiration on N_2O, may serve as *Anaeromyxobacter* adaptations to microaerophilic soil conditions (Thomas et al., 2010). Hence, further comparative and environmental genomics studies may be needed to elucidate the roles of versatile OHRB in mitigating the reactive by-products of denitrification.

2.5 Chapter 2 Figures.



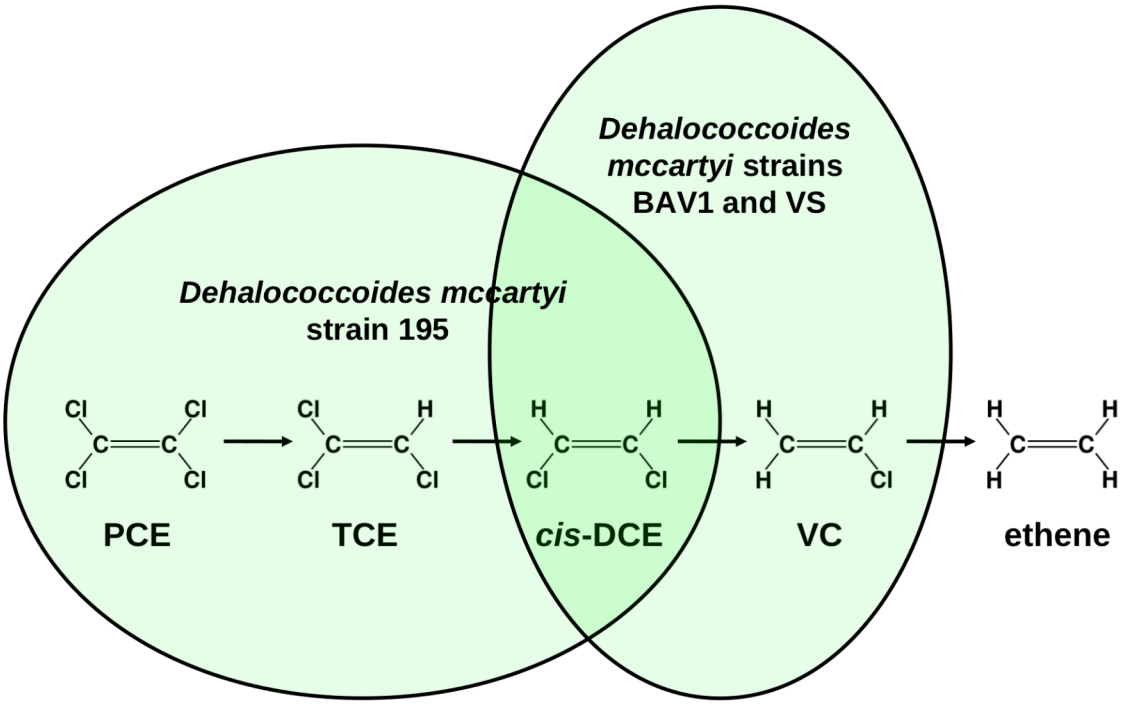
2.5.1 Figure 1. Physiology of organohalide respiration, A. Respiratory metabolism of an organohalide respiring bacterium. Anaerobic respiratory growth on an organohalide compounds exhibits three characteristics: i. A source of reducing equivalents such as H_2 or formate, denoted here as 2[H], ii. A membrane-associated respiratory complex, and iii. cleavage of at least one carbon-halide bond from carbon backbone of the organohalide substrate.

B. Electron flow through redox centers of reductive dehalogenase (RDase) complex. According to the scheme elucidated by Holliger et al. (1998) reducing equivalents available from H_2 or carbon source, denoted e-, reduce the iron-sulfur centers (4Fe-4S). The iron-sulfur centers in turn reduce the cobalt(III) of the corrinoid to cobalt(I), enabling reduction of the organohalide substrate. The RDase A subunit (green) contains both iron-sulfur and corrinoid redox centers while the RDase B subunit (yellow) is believed to function as a membrane anchor.



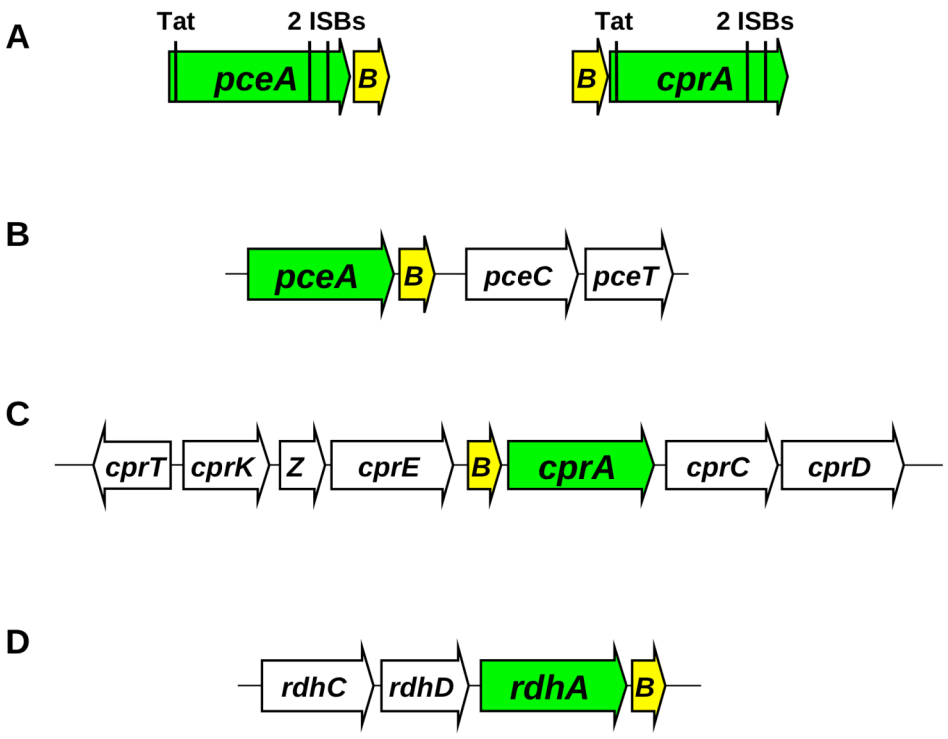
2.5.2 Figure 2. 16S rRNA gene phylogeny of isolated OHRB strains.

Representative OHRB strains from *ε-Proteobacteria, δ-Proteobacteria, Firmicutes,* and *Chloroflexi* are shown with type of organohalide substrate used during enrichment and isolation process. Black triangles denote growth and enrichment from chlorinated aliphatic substrates (e.g., PCE, TCE, vinyl chloride (VC)) while green hexagons denote growth and enrichment from chlorinated aromatics, such as chlorophenols or chlorobenzenes. Bootstrap support (out of 1000 replicates) are represented by half-closed circles (50% < support < 89%), and black circles (90% < support). The scale bar in the lower left corner corresponds to nucleotide changes at 10% of aligned sites.



2.5.3 Figure 3. Example of substrate range differences among OHRB.

Dehalococcoides mccartyi strains are collectively able to couple growth to the complete reduction of TCE and PCE to ethene. Differences in substrate utilization are related to differences in RDase gene/protein repertoire. However, no single OHRB isolate harbors the set of appropriate RDase enzymes to complete the pathway.



2.5.4 Figure 4. Representative RDase genes, A. RDase protein and gene sequence features. Gene organization and amino acid motif positions for PCE RDase (Suyama et al., 2002; Maillard et al., 2003) and chlorophenol RDase (van de Pas et al., 1999). **B. PCE RDase genes (*pceA-pceB*) and adjacent accessory genes.** The *pceA-B-C-T* cluster from genomes of *Desulfitobacterium hafniense* strains Y51 and TCE1 (Suyama et al., 2002; Maillard et al., 2005). **C. Chlorophenol RDase genes (*cprB-cprA*), accessory genes, and regulator, *cprK*.** The *cprT-K-Z-E-B-A-C-D* cluster on the genomes of *Desulfitobacterium dehalogenans* (Smidt, et al., 2000) and *Desulfitobacterium hafniense* strain DCB-2 (Kim et al., 2012) harbors the regulator gene, *cprK* (Pop et al., 2004). **D. Chlorobenzene RDase and several putative RDase gene loci and accessory genes.** *Dehalococcoides mccartyi* RDase genes often have organization *rdhA-B-C-D* or *rdhR-A-B*, where *rdhC, rdhD,* and *rdhR* encode regulators (Seshadri et al., 2005).

2.6 References.

Adrian, L., Dudková V., Demnerová K. and Bedard D. (2009). "*Dehalococcoides*" sp. strain CBDB1 extensively dechlorinates the commercial polychlorinated biphenyl mixture Arochlor 1260. *Applied and Environmental Microbiology* **75**: 4516-4524.

Adrian, L., Rahenführer J., Gobom J. and Hölscher T. (2007). Identification of a Chlorobenzene reductive dehalogenase in *Dehalococcoides* sp. Strain CBDB1. *Applied and Environmental Microbiology* **73**: 7717-7724.

Alexander, M. (1991). Research Needs in Bioremediation. *Environmental Science Technology* **25**: 1972 - 1973.

Altschul, S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Altschul, S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Banerjee, R. and Ragsdale S. W. (2003). The many faces of vitamin B_{12}: Catalysis by cobalamin-dependent enzymes. *Annual Reviews of Biochemistry* 72: 209-247.

Berks, B. C., Palmer T. and Sargent F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology* **8**: 174-181.

Bork, P. and Koonin E. V. (1996). Protein sequence motifs. *Current Opinion in Structural Biology* **6**: 366-376.

Bunge, M., Adrian L., Kraus A., Opel M., Lorenz W. G., Andreeson J. R., Görisch H. and Lechner U. (2003). Reductive dehalogenation of chlorinated dioxins by an anaerobic bacterium. *Nature* **421**: 357-360.

Cicerone, R. J. (1987). Changes in stratospheric ozone. *Science* **237**: 35-42.

Cole, J. R., Cascarelli A. L., Mohn W. W. and Tiedje J. M. (1994). Isolation and characterization of a novel bacterium growing via reductive dehalogenatioin of 2-chlorophenol. *Applied and Environmental Microbiology* **60**: 3536-3542.

Crutzen, P. J. (1970). Influence of nitrogen oxides on atmospheric ozone content. *Quarterly Journal of the Royal Meteorological Society* **96**: 320-325.

Dandekar, T., Snel B., Huynen M. and Bork P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *TIBS* **23**: 324-328.

DiStefano, T. D., Gossett J. M. and Zinder S. H. (1991). Reductive dechlorination of high concentrations of tetrachloroethene to ethene by an anaerobic enrichment culture in the absence of methanogenesis. *Applied and Environmental Microbiology* **57**: 2287-2297.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998) Biological sequence analysis: Probabalistic models of proteins and nucleic acids. *Cambridge University Press*.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Fantroussi, S. E., Naveau H. and Agathos S. N. (1998). Anaerobic dechlorinating bacteria. *Biotechnology Progress* **14**: 167-188.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics* **5**: 164-166.

Fetzner, S. and Lingens F. (1994). Bacterial dehalogenases: Biochemistry, genetics, and biotechnological applications. *Microbiological Reviews* **58**: 641-685.

Finlayson-Pitts, B. J. and Pitts J. N. (2000). *Chemistry of the Upper and Lower Atmosphere.* San Diego, Academic Press.

Finn, R.D., Mistry J., Tate J., Coggill P., Heger A., Pollington J.E., et al. (2010). The Pfam protein families database. *Nucleic Acids Research* **38**: 211-222.

Futagami, T., Yamaguchi T., Nakayama S.-i., Goto M. and Furukawa K. (2006). Effects of chloromethanes on growth of and deletion of the *pce* gene cluster in dehalorespiring *Desulfitobacterium hafniense* Strain Y51. *Applied and Environmental Microbiology* **72**: 5998-6003.

Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger S. P., et al. (2004). Nitrogen cycles: past, present, and future. *Biogeochemistry* **70**: 153-226.

Gantzer, C. J. and Wackett L. P. (1991). Reductive dechlorination catalyzed by bacterial transition-metal coenzymes. *Environmental Science Technology* **25**: 715 - 722.

Graedel, T. E. and Keene W. C. (1996). The budget and cycle of Earth's natural chlorine. Pure and Applied Chemistry 68: 1689-1697.

Gribble, G. W. (1998). Naturally Occurring Organohalogen Compounds. *Accounts of Chemical Research* **31**: 141-152.

Gribble, G. W. (2003). The diversity of naturally produced organohalogens. *Chemosphere* **52**: 289-297.

Häggblom, M. M., Ahn Y.-B., Fennell D. E., Kerkhof L. J. and Rhee S.-K. (2003). Anaerobic dehalogenation of organohalide contaminants in the marine environment. *Advances in Applied Microbiology* **53**: 61-84.

He, J., Holmes V. F., Lee P. K. H. and Alvarez-Cohen L. (2007). Influence of Vitamin B_{12} and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Applied and Environmental Microbiology* **73**: 2847-2853.

He, J., Ritalahti K. M., Yang K.-L., Koenigsberg S. S. and Löffler F. E. (2003). Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* **424**: 62-65.

Holliger, C., Hahn D., Harmsen H., Ludwig W., Schumacher W., Tindall B., Vazquez F., Weiss N. and Zehnder A. J. B. (1998). *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetrachloroethene and trichloroethene in an anaerobic respiration. *Archives of Microbiology* **169**: 313-321.

Holliger, C., Schraa G., Stams A. J. M. and Zehnder A. J. B. (1993). A highly purified enrichment culture couples the reductive dechlorination of tetrachloroethene to growth. *Applied and Environmental Microbiology* **59**: 2991-2997.

Holliger, C. and Schumacher W. (1994). Reductive dehalogenation as a respiratory process. *Antonie van Leeuwenhoek* **66**: 239-246.

Holliger, C., Wohlfarth G. and Diekert G. (1999). Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiology Reviews* **22**: 383-398.

Hölscher, T., Krajmalnik-Brown R., Ritalahti K. M., Wintzingerode F. v., Görisch H., Löffler F. E. and Adrian L. (2004). Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*. *Applied and Environmental Microbiology* **70**: 5290-5297.

John, M., Schmitz R. P. H., Westermann M., Richter W. and Diekert G. (2006). Growth substrate dependent localization of tetrachloroethene reductive dehalogenase in *Sulfurospirillum multivorans*. *Archives of Microbiology* 186: 99-106.

Johnson, B. L. (1995). Nature, extent, and impact of Superfund hazardous waste sites. *Chemosphere* **31**: 2415-2428.

Johnson, D. R., Nemir A., Andersen G. L., Zinder S. H. and Alvarez-Cohen L. (2009). Transcriptomic microarray analysis of corrinoid responsive genes in *Dehalococcoides ethenogenes* strain 195. *FEMS Microbiology Letters* **294**: 198-206.

Jordan, A., Harnisch J., Borchers R., Guern F. L. and Shinohara H. (2000). Volcanic Halocarbons. *Environmental Science and Technology* **34**: 1122-1124.

Kern, M. and Simon J. (2009). Electron transport chains and bioenergetics of respiratory nitrogen metabolism in *Wolinella succinogenes* and other Epsilonproteobacteria. *Biochimica et Biophysica Acta* **1787**: 646-656.

Kim, S.-H., Harzman C., Davis J. K., Hutcheson R., Broderick J. B., Marsh T. L. and Tiedje J. M. (2012). Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiology* **12**: 21.

Kliegman, S. and McNeill K. (2008). Dechlorination of chloroethylenes by cob(I)alamin and cobalamin model complexes. *Dalton Transactions* **32**: 4181-4320.

Koonin, E. V., Tatusov R. L. and Galperin M. Y. (1998). Beyond complete genomes: from sequence to structure to function. *Current Opinion in Structural Biology* **8**: 355-363.

Krajmalnik-Brown, R., Hölsher T., Thomson I. N., Saunders F. M., Ritalahti K. M. and Löffler F. E. (2004). Genetic identification of a putative vinyl chloride reductase in *Dehalococcoides* sp. Strain BAV1. *Applied and Environmental Microbiology* **70**: 6347-6351.

Krajmalnik-Brown, R., Sung Y., Ritalahti K. M., Saunders F. M. and Löffler F. E. (2007). Environmental distribution of the trichloroethene reductive dehalogenase gene (*tceA*) suggests lateral gene transfer among *Dehalococcoides*. *FEMS Microbiology Ecology* **59**: 206-214.

Krasotkina, J., Walters T., Maruya K. A. and Ragsdale S. W. (2001). Characterization of the B_{12}- and iron-sulfur-containing reductive dehalogenase from *Desulfitobacterium chlororespirans*. *The Journal of Biological Chemistry* **276**: 40991-40997.

Kraütler, B., Fieber W., Ostermann S., Fasching M. and Ongania K.-H. (2003). The cofactor of tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans* is Norpseudo-B_{12} , a new type of a natural corrinoid. *Helvetica Chimica Acta* **86**: 3698-3716.

Kube, M., Beck A., Zinder S. H., Kuhl H., Reinhardt R. and Adrian L. (2005). Genome sequence of the chlorinated compound-respiring bacterium *Dehalococcoides* species strain CBDB1. *Nature Biotechnology* **23**: 1269-1273.

Lashof, D. A. and Ahuja D. R. (1990). Relative contributions of greenhouse gas emissions to global warming. *Nature* **344**: 529-531.

Liu, X., Gao C., Zhang A., Jin P., Wang L. and Feng L. (2008). The *nos* gene cluster from gram-positive bacterium *Geobacillus thermodenitrificans* NG80-2 and functional characterization of the recombinant NosZ. *FEMS Microbiology Letters* **289**: 46-52.

Löffler, F. E. and Edwards E. A. (2006). Harnessing microbial activities for environmental cleanup. *Current Opinion in Biotechnology* **17**: 274-284.

Löffler, F. E., Yan J., Ritalahti K. M., Adrian L., Edwards E. A., Konstantinidis K. T., et al. (2012). *Dehalococcoides mccartyi* gen. nov., sp. nov., obligate organohalide-respiring anaerobic bacteria, relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidetes* classis nov., within the phylum *Chloroflexi*. *International Journal of Systematic and Evolutionary Microbiology* **Apr 27** Epub ahead of print.

Louie, T. M. and Mohn W. W. (1999). Evidence for a chemiosmotic model of dehalorespiration in *Desulfomonile tiedjei* DCB-1. *Journal of Bacteriology* **181**: 40-46.

Lovley, D. R. (2006). Bug juice: harvesting electricity with microorganisms. *Nature Reviews* **4**: 497-508.

Ludwig, M. L. and Matthews R. G. (1997). Structure-based perspectives on B_{12}-dependent enzymes. *Annual Review of Biochemistry* **66**: 269-313.

Madsen, T. and Licht D. (1992). Isolation and characterization of an anaerobic chlorophenol-transforming bacterium. *Applied and Environmental Microbiology* **58**: 2874-2878.

Magnuson, J. K., Stern R. V., Gosset J. M., Zinder S. H. and Burris D. R. (1998). Reductive dechlorination of tetrachloroethene to ethane by a two-component enzyme pathway. *Applied and Environmental Microbiology* **64**: 1270-1275.

Maillard, J., Schumacher W., Vazquez F., Regeard C., Hagen W. R. and Holliger C. (2003). Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus*. *Applied and Environmental Microbiology* **69**: 4628-4638.

Manassaram, D. M., Backer L. C. and Moll D. M. (2006). A review of nitrates in drinking water: Maternal exposure and adverse reproductive and developmental outcomes. *Environmental Health Perspectives* **114**: 320-327.

Maphosa, F., Vox W. M. d. and Smidt H. (2010). Exploiting the ecogenomics toolbox for environmental diagnostics of organohalide-respiring bacteria. *Trends in Biotechnology* **28**: 308-316.

Marietou, A., Richardson D., Cole J. and Mohan S. (2005). Nitrate reduction by *Desulfovibrio desulfuricans*: a periplasmic nitrate reductase system that lacks NapB, but includes a unique tetraheme *c*-type cytochrome, NapM. *FEMS Microbiology Letters* 248(2): 217-225.

Maymó-Gatell, X., Anguish T. and Zinder S. H. (1999). Reductive dechlorination of chlorinated ethenes and 1,2-dichloroethane *Applied and Environmental Microbiology* **65**: 3108-3113.

Maymó-Gatell, X., Chien Y.-t., Gosset J. M. and Zinder S. H. (1997). Isolation of a bacterium that reductively dechlorinates tetrachlorethene to ethene. *Science* **276**: 1568-1571.

McLain, J. E. T. and Martens D. A. (2005). Nitrous oxide flux from soil amino acid mineralization. *Soil Biology and Biochemistry* **37**: 289-299.

McMurdie, P. J., Behrens S., Müller J. A., Göke J., Ritalahti K. M., Wagner R., et al. (2009). Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genetics* **5**: e1000714.

McMurdie, P. J., Behrens S. F., Holmes S. and Spormann A. M. (2007). Unusual codon bias in vinyl chloride reductase genes of *Dehalococcoides* species. *Applied and Environmental Microbiology* **73**: 2744-2747.

Men, Y., Feil H., VerBerkmoes N. C., Shah M. B., Johnson D. R., Lee P. K. H., et al. (2012). Sustainable syntrophic growth of *Dehalococcoides ethenogenes* strain 195 with *Desulfovibrio vulgaris* Hildenborough and *Methanobacterium congolense*: global transcriptomic and proteomic analyses. *The ISME Journal* **6**: 410-421.

Miller, E., Wohlfarth G. and Diekert G. (1997). Comparative studies on tetrachloroethene reductive dechlorination mediated by *Desulfitobacterium* sp. strain PCE-S. *Archives of Microbiology* **168**: 513-519.

Miller, E., Wohlfarth G. and Diekert G. (1998). Purification and characterization of the tetrachloroethene reductive dehalogenase of strain PCE-S. *Archives of Microbiology* **169**: 497-402.

Mohn, W. W. and Kennedy K. J. (1992). Reductive dehalogenation of chlorophenols by *Desulfomonile tiedjei* DCB-1. *Applied and Environmental Microbiology*. **58**: 1367-1370.

Morita, Y., Futagami T., Goto M. and Furukawa K. (2009). Functional characterization of the trigger factor protein PceT of tetrachloroethene-dechlorinating *Desulfitobacterium hafniense* Y51. *Applied Microbiology and Biotechnology* 83: 775-781.

Morris, R. M., Fung J. M., Rahm B. G., Zhang S., Freedman D. L., Zinder S. H. and Richardson R. E. (2007). Comparative proteomics of *Dehalococcoides* spp. reveals strain-specific peptides associated with activity. A*pplied and Environmental Microbiology* 73(1).

Morris, R. M., Sowell S., Barofsky D., Zinder S. H. and Richardson R. E. (2006). Transcription and mass-spectroscopic proteomic studies of electron transport oxidoreductases in *Dehalococcoides ethenogenes. Environmental Microbiology* **8**.

Müller, J. A., R. B. M., Abendroth G. v., Meshulam-Simon G., McCarty P. L. and Spormann A. M. (2004). Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. Strain VS and its environmental distribution. *Applied and Environmental Microbiology* **70**: 4880-4888.

Neumann, A., Wohlfarth G. and Diekert G. (1996). Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*. The *Journal of Biological Chemistry* **271**: 16515-16519.

Ni, S., Fredrickson J. K. and Xun L. (1995). Purification and characterization of a novel 3-chlorobenzoate-reductive dehalogenase from the cytoplasmic membrane of *Desulfomonile tiedjei* DCB-1. *Journal of Bacteriology* **177**: 5135-5139.

Nijenhuis, I. and Zinder S. H. (2005). Characterization of hydrogenase and reductive dehalogenase activities of *Dehalococcoides ethenogenes* Strain 195. *Applied and Environmental Microbiology* **71**: 1664-1667.

Nolan, B. T. and Ruddy B. C. (1996). Nitrate in the Ground Waters of the United States--Assessing the Risk. U. S. G. Survey. Reston, VA. FS-092-96.

Nonaka, H., Keresztes G., Shinoda Y., Ikenaga Y., Abe M., Naito K., et al. (2006). Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195. *Journal of Bacteriology* **188**: 2262-2274.

Nowak, J., Kirsch N. H., Hegemann W. and Stan H. J. (1996). Total reductive dechlorination of chlorobenzenes to benzene by a methanogenic mixed culture enriched from Saale river sediment. *Applied Microbiology and Biotechnology* **45**: 700-709.

Öberg, G. (1998). Chloride and organic chlorine in soil. *Acta hydrochimica hydrobiology* **26**: 137-144.

Park, J., Karplus K., Barrett C., Hughey R., Haussler D., Hubbard T. and Chothia C. (1998). Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *Journal of Molecular Biology* **284**: 1201-1210.

Payne, W. J., Grant M. A., Shapleigh J. and Hoffman P. (1982). Nitrogen oxide reduction in *Wolinella succinogenes* and *Campylobacter* species. *Journal of Bacteriology* **152**: 915-918.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucleic Acids Research* **35**: 61-65.

Philippot, L. (2002). Denitrifying genes in bacterial and Archaeal genomes. *Biochimica et Biophysica Acta* **1577**: 355-376.

Pohl, H. R., Tarkowski S., Buczynska A., Fay M. and Rosa C. T. D. (2008). Chemical exposures at hazardous waste sites: Experiences from the United States and Poland. *Environmental Toxicology and Pharmacology* **25**: 283-291.

Pop, S. M., Kolarik R. J. and Ragsdale S. W. (2004). Regulation of anaerobic dehalorespiration by the transcriptional activator CprK. *The Journal of Biological Chemistry* **279**: 49910-49918.

Rahm, B. G., Morris R. M. and Richardson R. E. (2006). Temporal expression of respiratory genes in an enrichment culture containing *Dehalococcoides ethenogenes. Applied and Environmental Microbiology* **72**: 5486-5491.

Ravishankara, A. R., Daniel J. S. and Portmann R. W. (2009). Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. *Science* **326**: 123-125.

Reeck, G. R., Haën C. d., Teller D. C., Doolittle R. F., Fitch W. M., Dickerson R. E., et al. (1987). "Homology" in proteins and nucleic acids: A terminology muddle and a way out of it. *Cell* **50**: 667.

Rosner, B. M., McCarty P. L. and Spormann A. M. (1997). In vitro studies on reductive vinyl chloride dehalogenation by an anaerobic mixed culture. Applied and Environmental Microbiology 63(11): 4139-4144.

Sanford, R. A., Cole J. R., Löffler F. E. and Tiedje J. M. (1996). Characterization of *Desulfitobacterium chlororespirans* sp. nov., which grows by coupling the oxidation of lactate to the reductive dechlorination of 3-chloro-4-hydroxybenzoate. *Applied and Environmental Microbiology* **62**: 3800-3808.

Sanford, R. A., Cole J. R. and Tiedje J. M. (2002). Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68**: 893-900.

Scholz-Muramatsu, H., Neumann A., Messmer M., Moore E. and Diekert G. (1995). Isolation and characterization of *Dehalospirillum multivorans* gen. nov. sp. nov., a tetrachloroethene-utilizing, strictly anaerobic bacterium. *Archives of Microbiology* **163**: 48-56.

Schumacher, W. and Holliger C. (1996). The proton/electron ratio of the menaquinone-dependent electron transport from dihydrogen to tetrachloroethene in "*Dehalobacter restrictus*". *Journal of Bacteriology* **178**: 2328-2333.

Schumacher, W., Holliger C., Zehnder A. J. B. and Hagen W. R. (1997). Redox chemistry of cobalamin and iron-sulfur cofactors in the tetrachloroethene reductase of *Dehalobacter restrictus. FEBS Letters* **409**: 421-425.

Scott, A. I. and Roessner C. A. (2002). Biosynthesis of cobalamin (vitamin B_{12}). *Biochemical Society Transactions* **30**: 613-620.

Seshadri, R., Adrian L., Fouts D. E., Eisen J. A., Phillippy A. M., Methe B. A., et al. (2005). Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105-108.

Shelton, D. R. and Tiedje J. M. (1984). Isolation and partial characterization of bacteria in an anaerobic consortium that mineralizes 3-chlorobenzoic acid. *Applied and Environmental Microbiology* **48**: 840-848.

Siebert, A., Neumann A., Schubert T. and Diekert G. (2002). A non-dechlorinating strain of *Dehalospirillum multivorans*: evidence for a key role of the corrinoid cofactor in the synthesis of an active tetrachloroethene dehalogenase. *Archives of Microbiology* **178**: 443-449.

Simon, J., Einsle O., Kroneck P. M. H. and Zumft W. (2004). The unprecedented *nos* gene cluster of *Wolinella succinogenes* encodes a novel respiratory electron transfer pathway to cytochrome *c* nitrous oxide reductase. *FEBS Letters* 569: 7-12.

Smidt, H., Akkermans A. D. L., Oost J. v. d. and Vos W. M. d. (2000). Halorespiring bacteria - molecular characterization and detection. *Enzyme and Microbial Technology* **27**: 812-820.

Smidt, H., Leest M. v., Oost J. v. d. and Vos W. M. d. (2000). Transcriptional regulation of the cpr gene cluster in ortho-chlorophenol-respiring *Desulfitobacterium dehalogenans*. *Journal of Bacteriology* **182**: 5683-5691.

Smidt, H. and Vos W. M. d. (2004). Anaerobic microbial dehalogenation. *Annual Review of Microbiology* **58**: 43-73.

Squillace, P. J., Scott J. C., Moran M. J., Nolan B. T. and Kolpin D. W. (2002). VOCs, pesticides, nitrate, and their mixtures in groundwater used for drinking water in the United States. *Environmental Science and Technology* **36**: 1923-1930.

Stein, L. Y. and Yung Y. L. (2003). Production, isotopic composition, and atmospheric fate of biologically produced nitrous oxide. *Annual Review of Earth and Planetary Sciences* **31**: 329-356.

Strycharz, S. M., Gannon S. M., Boles A. R., Franks A. E., Nevin K. P. and Lovley D. R. (2010). Reductive dechlorination of 2-chlorophenol by *Anaeromyxobacter dehalogenans* with an electrode serving as the electron donor. *Environmental Microbiology Reports* **2**: 289-294.

Strycharz, S. M., Woodard T. L., Johnson J. P., Nevin K. P., Sanford R. A., Löffler F. E. and Lovley D. R. (2008). Graphite electrode as a sole electron donor for reductive dechlorination of tetrachloroethene by *Geobacter lovleyi*. *Applied and Environmental Microbiology* **74**: 5943-5947.

Sung, Y., Fletcher K. E., Ritalahti K. M., Apkarian R. P., Ramos-Hernández N., Sanford R. A., Mesba N. M. and Löffler F. E. (2006a). *Geobacter lovleyi* sp. nov. strain SZ, a novel metal-reducing and tetrachloroethene-dechlorinating bacterium. *Applied and Environmental Microbiology* **69**: 2775-2782.

Sung, Y., Ritalahti K. M., Apkarian R. P. and Löffler F. E. (2006b). Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring *Dehalococcoides* isolate. *Applied and Environmental Microbiology* **72**: 1980-1987.

Sung, Y., Ritalahti K. M., Sanford R. A., Urbance J. W., Flynn S. J., Tiedje J. M. and Loffler F. E. (2003). Characterization of two tetrachloroethene-reducing, acetate-oxidizing anaerobic bacteria and their description as *Desulfuromonas michiganensis* sp. nov. *Applied and Environmental Microbiology* **69**: 2964-2974.

Suyama, A., Iwakiri R., Kai K., Tokunaga T., Sera N. and Furukawa K. (2001). Isolation and characterization of *Desulfitobacterium* sp. strain Y51 capable of efficient dehalogenation of tetrachloroethene and polychloroethanes. *Bioscience, Biotechnology, and Biochemistry* **65**: 1474-1481.

Suyama, A., Yamashita M., Yoshino S. and Furukawa K. (2002). Molecular characterization of the PceA reductive dehalogenase of *Desulfitobacterium* sp. Strain Y51. *Journal of Bacteriology* **184**: 3419-3425.

Tandol, V., DiStefano T. D., Bowser P. A., Gossett J. M. and Zinder S. M. (1994). Reductive dehalogenation of chlorinated ethenes and halogenated ethanes by a high-rate anaerobic enrichment culture. *Environmental Science Technology* **28**: 973 - 979.

Tas, N., Eekert M. H. A. v., Vos W. M. d. and Smidt H. (2010). The little bacteria that can - diversity, genomics, and ecophysiology of '*Dehalococcoides*' spp. in contaminated environments. *Microbial Biotechnology* **3**: 389-402.

Thomas, S. H., Sanford R. A., Amos B. K., Leigh M. B., Cardenas E. and Löffler F. E. (2010). Unique ecophysiology among U(VI)-reducing bacteria as revealed by evaluation of oxygen metabolism in *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology* **76**: 176-183.

Thomas, S. H., Wagner R. D., Arakaki A. K., Skolnick J., Kirby J. R., Shimkets L. J., Sanford R. A. and Löffler F. E. (2008). The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* **3**: e2103.

Utkin, I., Woese C. and Wiegel J. (1994). Isolation and characterization of *Desulfitobacterium dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *International Journal of Systematic Bacteriology* **44**: 612-619.

Vallack, H. W., Bakker, D. J., Brandt, I., Broström-Lundén, E., Brouwer A., Bull K. R., et al. (1998). Controlling persistent organic pollutants - what next? *Environmental Toxicology and Pharmacology* **6**: 143-175.

van den Pas, B. A., Jansen S., Dijkema C., Schraa G., Vos W. M. d. and Stams A. J. M. (2001). Energy yield of respiration on chloroaromatic compounds in *Desulfitobacterium dehalogenans*. *Applied and Environmental Microbiology* **67**: 3958-3963.

van den Pas, B. A., Smidt H., Hagen W. R., Oost J. v. d., Schraa G., Stams A. J. M. and Vos W. M. d. (1999). Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in *Desulfitobacterium dehalogenans*. *The Journal of Biological Chemistry* **274**.

van Eekert, M. H. A. and Schraa G. (2001). The potential of anaerobic bacteria to degrade chlorinated compounds. *Water Science and Technology* **44**: 49-56.

Vetter, W. and Gribble, G. (2007). Anthropogenic persistent organic pollutants - lessons to learn from halogenated natural products. *Environmental Toxicology and Chemistry* **26**: 2249-2252.

Vogel, T. M., Criddle C. S. and McCarty P. L. (1987). Transformations of halogenated aliphatic compounds. *Environmental Science Technology* **21**: 722 - 736.

Wagner, D. D., Hug L. A., Hatt J. K., Spitzmiller M. R., Padilla-Crespo E., Ritalahti K. M., Edwards E. A., Konstantinidis K. T. and Löffler F. E. (2012). Genomic determinants of organohalide-respiration in *Geobacter lovleyi*, an unusual member of the *Geobacteraceae*. *BMC Genomics* **13**: 200.

Wang, W. C., Yung Y. L., Lacis A. A., Mo T. and Hansen J. E. (1976). Greenhouse effects due to man-made perturbations of trace gases. *Science* **194**: 685-690.

Warren, M. J., Raux E., Schubert H. L. and Escalante-Semerena J. C. (2002). The biosynthesis of adenosylcobalamin (vitamin B_{12}). *Natural Product Reports* **19**: 390-412.

Winterton, N. (2000). Chlorine: the only green element - towards a wider acceptance of its role in natural cycles. *The Royal Society of Chemistry* **2**: 173-225.

Yan, J., Ritalahti K. M., Wagner D. D. and Löffler F. E. (2012). Unexpected specificity of interspecies cobamide transfer from *Geobacter* spp. to organohalide-respiring *Dehalococcoides mccartyi* strains. *Applied and Environmental Microbiology* **78**: 6630-6636.

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews* 61(4): 533-616.

Zumft, W. G. and Körner H. (2007). *Nitrous oxide reductases. Biology of the nitrogen cycle*. H. Bothe, S. J. Ferguson and W. E. Newton. Amsterdam, Elsevier: 67-81.

Zumft, W. G. and Kroneck P. M. H., Eds. (2007). Respiratory transformation of nitrous oxide (N_2O) to dinitrogen by Bacteria and Archaea. *Advances in Microbial Physiology*. London, Academic Press Ltd.-Elsevier Science Ltd.

CHAPTER 3

EVOLUTIONARY HISTORY AND FUNCTIONAL DIFFERENTIATION OF REDUCTIVE DEHALOGENASES FROM ORGANOHALIDE-RESPIRING *BACTERIA.*

3.1 Abstract.

Reductive dehalogenases (RDases) of organohalide-respiring *Bacteria* (OHRB) are membrane-associated oxidoreductases mediating reduction of chloroorganic electron acceptors. Sequence features serving to identify catalytic reductive dehalogenase A (RDase A) proteins are not clearly defined, complicating functional prediction and genome annotation. A hidden Markov model based upon 12 validated RDase A proteins identified 406 putative RDases, designated class 1, shown by manual curation to possess N-terminal Tat- or Sec-type signal peptides plus associated B proteins allowing interaction with the membrane. N-terminal Sec-type signals and a conserved membrane topology was predicted for 190 complete RDase B proteins. The RDase HMM also retrieved 59 sequences, designated class 2, lacking both the N-terminal Tat consensus sequence and the associated RDase B gene, indicating cytoplasmic localization. Multiple sequence alignment (MSA) and HMM analyses identified conserved residues and domain-like regions shared between class 1 and class 2, suggesting the latter is part of the same superfamily, presumably with non-respiratory RDase function. Bootstrapped phylogenetic trees divided the RDase superfamily into 25 evolutionary lineages (i.e., clades), 14 comprising class 1 and 11 comprising class 2 sequences, and suggested polyphyly. Constrained phylogenetic analyses corroborated that class 1 RDase A proteins have polyphyletic origins. The analysis further suggested that RDase A proteins with overlapping substrate specificities are of polyphyletic origin, presumably due to convergent evolution. Three horizontal gene transfer (HGT) events amongst genera were inferred for 218 class 1 sequences, and no HGT events occurred at higher taxonomic levels. The deep phylogenetic branching, polyphyly, and evidence for HGT events suggest a long evolutionary history of RDase genes prior to recent releases of organohalide contaminants.

3.2 Introduction.

Organohalide respiring *Bacteria* (OHRB) representing *Firmicutes* (low GC Gram-positive *Bacteria*), *Chloroflexi*, ε-*Proteobacteria*, and δ-*Proteobacteria* have been isolated from a variety of anoxic environments including soils and aquifers contaminated with chlorinated ethenes (Holliger et al., 1993; Neumann et al., 1996; Holliger et al., 1998; Sanford et al., 2002; He et al., 2003; Villemur et al., 2006), aquatic sediments (Häggblom et al., 2003; Sung et al., 2003; Sung et al., 2006b), and activated sludge (Neumann et al., 1994; Scholz-Muramatsu et al., 1995; Maymó-Gatell et al. 1997). RDases of OHRB are respiratory terminal reductases involved in detoxification of contaminants such as chlorinated aliphatic compounds (e.g., chlorinated solvents) (Löffler and Edwards, 2006) and chlorinated aromatic compounds, such as chlorobenzenes (CBs) and chlorophenols (CPs) (Smidt and de Vos, 2004). At present, only 16 RDases have been characterized by (partial) enzyme purification or expression studies (Table S1) and none has structural information. Tetrachloroethene (PCE) RDases (PceA, EC:1.97.1.8) have been characterized from the ε-Proteobacterium *Sulfurospirillum multivorans* (Neumann et al., 1996; Neumann et al., 1998; Neumann et al., 2002), the low GC Gram-positive *Desulfitobacterium hafniense* strain PCE-S and strain Y51 (Miller et al., 1998; Suyama et al., 2002; Furukawa et al., 2005) and *Dehalobacter restrictus* strain PER-K23 (Schumacher et al., 1997; Maillard et al., 2003), and *Dehalococcoides mccartyi* strain 195, a member of the *Chloroflexi* (Magnuson et al., 1998). Three chlorophenol/chlorophenylacetate reductases (CprA) have been characterized from *Desulfitobacterium* spp. (Löffler et al., 1996; Christiansen et al., 1998; van de Pas et al., 1999; Krasotkina et al., 2001, Thibodeau et al., 2004), and a chlorobenzoate reductase has been described from the-

δ-Proteobacterium *Desulfomonile tiedjei* (Ni et al., 1995). Additional RDases are known among the *Dehalococcoidetes*, such as the TCE reductase (TceA) from *Dehalococcoides mccartyi* strain 195 (Magnuson et al., 1998; Magnuson et al., 2000), the chlorinated benzene reductase (CbrA) from *Dehalococcoides mccartyi* strain CBDB1 (Adrian et al., 2007), and the vinyl chloride (VC) reductase BvcA from *Dehalococcoides mccartyi* strain BAV1 (Krajmalnik-Brown et al., 2004) and VcrA from strain VS (Müller et al., 2004) and strain GT (Sung et al. 2006a). With the possible exception of the *Desulfomonile tiedjei* 3-chlorobenzoate reductase, dechlorination activities of known RDases depend on a corrinoid cofactor and two 4Fe-4S (or 3Fe-4S) iron-sulfur clusters (Miller et al., 1998; van den Pas et al., 1999; Krasotkina et al., 2001; Siebert et al., 2002; Kräutler et al., 2003; Maillard et al., 2003).

Sequencing of OHRB genomes generated an extensive pool of putative RDase genes encoding proteins with predicted similarity to biochemically characterized RDases (Table S2). In particular, the genomes of dechlorinating *Chloroflexi* harbor large numbers of putative RDase genes, and *Dehalococcoides mccartyi* strain 195 encodes 17, strain CBDB1 32 (Kube et al., 2005), strain GT 20, strain BAV1 11, and strain VS 36 RDases (McMurdie et al., 2009; Löffler et al. 2012). Other OHRB genomes, such as those of the *δ-Proteobacteria Geobacter lovleyi* strain SZ and *Anaeromyxobacter dehalogenans* strain 2CP-C, which respire PCE (Sung et al., 2006b) and 2-chlorophenol (Cole et al., 1994), respectively, encode two putative RDases each. All putative RDases share two C-terminal iron-sulfur binding (ISB) motifs, **FC**xx**C**xx**C**xxx**CP** and **C**x_n**C**x_2**C**xxx**C**, with characterized RDases (Hölscher et al., 2004; Smidt and Vos, 2004), but it is not clear if the ISB motifs discriminate RDases from non-RDase iron sulfur proteins. All functionally characterized RDase

sequences possess the N-terminal amino acid motif **RR**x**F** characteristic of the twin-arginine translocation (Tat) signal (Berks et al., 2005). Amino acid regions resembling the cobalamin-binding motif **D**xx**H**xx**G**-x_n-**S**x**L**-x_n-**GG** occur in only nine putative RDases (Ludwig and Matthews, 1997; Hölscher et al., 2004; Futamata et al., 2009) but are found in none of the characterized RDases (Smidt and de Vos, 2004). In addition, gene loci encoding the experimentally characterized RDases (e.g., *pceA*, Fig. 1) have an adjacent gene encoding a B protein ranging 72 – 104 amino acids in length with 2 - 3 predicted membrane-spanning helices suggesting function as a membrane anchor for the catalytic RDase A (van den Pas et al., 1999; Magnuson et al., 2000; Suyama et al., 2002; Maillard et al., 2003). Similarly, genes encoding putative RDases on *Chloroflexi* genomes (Kube et al., 2005; Seshadri et al., 2005; McMurdie et al., 2009) are generally associated with a B gene encoding a protein with a characteristic **WY**x**W** amino acid motif (Hölscher et al., 2004; Krajmalnik-Brown et al., 2004). Despite such common features among RDases associated with organohalide respiration, criteria employed in automated genome annotation of putative RDases are not clearly defined. For instance, it is not clear if RDases lacking an associated B gene or the Tat translocation signature are functional. Furthermore, predictions of RDase protein structural features relevant to respiratory function, corrinoid binding, or substrate specificity are lacking.

The current study employs Hidden Markov Models (HMMs) (Eddy, 1998) built from multiple sequence alignments (MSAs) of 12 characterized RDases (Fig. S1, Table S1) to identify shared protein sequences. HMMs also provide a means to identify structural domains (Eddy, 1998), sequence features unique to RDases, and possibly conserved amino acid residues involved in catalysis (Todd et al., 2001). The integrated analyses of HMMs, MSAs, phylogenetic trees, and genome sequences

identified novel shared RDase features, elucidated RDase evolutionary histories, and promises to improve genome sequence annotation.

3.3 Results.

3.3.1 HMM-based prediction of protein sequences related to RDases.

RDase HMM database searches were combined with manual curation of retrieved sequences to identify proteins evolutionarily related to RDases (Fig. 1). The HMM retrieved 1,450 sequences from GenBank, RefSeq, Interpro, and the draft genomes of *Dehalobacter* sp. strain CF-50 and Candidatus Dehalobium chlorocoercia strain DF-1, excluding sequences of less than 200 amino acids in length. The presence of an N-terminal Tat signal or linkage with a B gene encoding a predicted membrane-anchor B protein (Fig. 1) validated the relatedness of the retrieved sequences with characterized RDases. Tat or Sec-type signal peptides were identified in 256 sequences. Gene loci encoding 220 of the Tat-containing proteins were associated with membrane-anchor protein genes. An additional 150 proteins were from incomplete ORFs with insufficient N-terminal residues to confirm a Tat signal, yet were linked with genes encoding membrane-anchor proteins (Table S3) or shared >70% amino acid identity with retrieved sequences possessing the N-terminal Tat motif. In total, 406 sequences retrieved by the HMM with bit scores > 157 were shown by subsequent curation steps to possess the Tat or Sec signals, an adjacent B gene, or both, and were designated class 1 (Fig. 1; Table S3).

The application of a 110 bit score threshold of log-odds similarity with the HMM identified 59 additional sequences; however, these sequences lacked evidence of N-terminal signal peptides or associated membrane anchor proteins. Fifty-nine such sequences with bit scores ranging from 110 to 290 lacked functionally or

structurally characterized representatives, and were designated class 2 (Fig. 1, Table S4). Forty-three of the class 2 sequences were retrieved by the HMM within the same range of bit scores as class 1 sequences (bit scores > 157). The remaining 982 sequences ranked below all class 1 and class 2 sequences with bit scores < 110 (Fig. 1), and were termed "unclassified iron-sulfur proteins". The 982 low ranked sequences included experimentally characterized polyferredoxins and electron transfer iron-sulfur proteins from *Archaea* and *Bacteria* incapable of organohalide respiration (Table S5, A and B). Non-RDase HMMs constructed from the polyferredoxins and electron transfer proteins retrieved 786 sequences with bit scores > 100 from among the 982 low-ranked sequences. Classes 1 and 2 sequences were also retrieved by the non-RDase HMMs, but with bit scores < 50. These analyses suggested that the bit score threshold of 110 (e-value = 2.5e-24) is a sufficiently stringent threshold for inferring evolutionary relatedness with RDases.

3.3.2 Genomic affiliation of class 1 and class 2 sequences.

Differences in genomic affiliations for class 1 and class 2 sequences suggest divergent function; however, protein sequence similarities between characterized class 1 RDases and class 2 proteins suggest evolutionary relatedness. Of the 406 class 1 sequences, less than half (183) represented characterized RDase A proteins or putative RDase A genes identified on complete or draft genomes of OHRB (Table 1; Table S2-S3), and 210 were derived from pure (Table 1) or mixed cultures (Table S6) with demonstrated organohalide respiration activity. An additional 13 class 1 sequences were encoded on genomes lacking experimental evidence for organohalide respiration, all but two from genomes of marine *γ-Proteobacteria* (Table 2). Altogether, 97% of class 1 sequences originated from genomes or cultures associated with organohalide respiration. Given

their predicted localization in the cytoplasm, class 2 proteins may not function in membrane-associated organohalide respiration. Accordingly, out of the 59 class 2 sequences, only three were affiliated with characterized OHRB, Dhaf_0693 from *Desulfitobacterium hafniense* strain DCB-2 and Desde_4096 and Desde_0044 from *Desulfitobacterium dehalogenans*. Twenty-five class 2 sequences were from genomes of marine α- and γ-*Proteobacteria* lacking evidence of organohalide respiration activity (Table 3). These marine *Proteobacteria* class 2 sequences exhibited three additional PFAM domains C-terminal to the two ISB sites, a 2Fe-2S domain, two flavin-binding domains (http://janelia.pfam.org), and ranged in length from 1,047 to 1,102 amino acids (Fig. 1, center), contrasting with the class 1 sequence lengths ranging from 310 to 554 amino acids (avg. = 488). Other class 2 sequences were found on the genomes of the extremophilic *Firmicutes Dethiobacter alkaliphilus* and *Heliobacterium modesticaldum*, the *Crenarchaeota Ferroglobus placidus*, and pathogenic *Clostridium difficile* strains (Table 2), with protein lengths ranging from 345 to 407 amino acids (Fig. 1, bottom). Despite absence of sequence features associated with respiratory RDases and affiliation with organisms not shown to be capable of reductive dehalogenation, class 2 sequences share Blast matches with functionally characterized RDases. For example, the class 2 sequence SPO0589 from *Ruegeria pomeroyi* strain DSS-3 shares 42% amino acid identity (57% similarity, ignoring length differences) in Blast with CprA (Dhaf_0713) from *Desulfitobacterium hafniense* strain DCB-2. The class 2 sequence HM1_1546 from *Heliobacterium modesticaldum* strain Ice1 shared 40% amino acid identity (56% similarity) in Blast with the biochemically characterized PceA of *Sulfurospirillum multivorans* (AAC60788). Hence, genomic affiliations of class 1 and class 2 sequences suggest

broader taxonomic and environmental distribution of proteins similar to RDases than is known for experimentally characterized RDases.

3.3.3 Amino acid sequence features reveal evolutionary relatedness exists between class 1 and class 2 proteins.

MSAs and custom sliding window HMMs identified protein features conserved across membrane-associated class 1 and cytoplasmic class 2 proteins, suggesting both classes may comprise the same superfamily. Inclusion of unclassified sequences in MSA and HMM analyses revealed that amino acid motifs shared among class 1 and class 2 proteins were not conserved in unclassified iron-sulfur proteins. Similarities and differences in site conservation across 401 out of 406 class 1, 57 out of 59 class 2, and nine unclassified iron-sulfur proteins comprising the MSA were apparent as variations in amino acid residues conserved in 80% to 90% of aligned sequences (i.e., consensus sequences, Table 3). The first conserved region, **PE**x_{13}**G**x_5**G**, located in the center of the 467-sequence MSA exhibited an 80% consensus (Fig. 2, top center). The proline residue was lacking in each of the nine unclassified iron-sulfur proteins. In the C-terminal half of the 467-sequence MSA, consensus motifs were identified that separated RDase superfamily class 1 and class 2 proteins from unclassified iron-sulfur proteins (Table 3). The 401 class 1 sequences shared the 90% consensus motif **Y**x_{12}**F**x_3**LGY** (Fig. 2, top center). At the same sites, 57 class 2 sequences had the variant 90% consensus motif **Y**x_{13}**IR**x_2**GY**. An adjacent C-terminal region exhibited the 90% consensus motif **G**xx**E**xx**R** across both class 1 and class 2 proteins (Fig. 2, top center). The nine unclassified iron-sulfur proteins possessed only the conserved glycine residues in the consensus sequences **Y**x_{12}**(F** or **IR)**x_n**LGY** and **G**xx**E**xx**R**. The two ISB sites **FC**$x_{(2-3)}$**C**xx**C**xxx**C**$x_{(0-8)}$**P** and **C**$x_{(2-21)}$**C**$x_{(2-9)}$**C**xxx**C**x**F** were conserved

with 90% consensus across class 1, class 2, and unclassified iron-sulfur proteins (Fig. 2, lower right). C-terminal to the ISB sites, a histidine (H) and a tyrosine (Y) residue, each with 90% consensus among class 1 sequences only (Fig. 2, top right). Conservation at multiple sites of consensus motifs unique to class 1 and class 2 proteins suggest membership in the superfamily despite apparent differences in class 1 and class 2 function.

Customized sliding window RDase HMMs provided evidence that class 1 and class 2 sequences shared highly-conserved central and C-terminal amino acid regions lacking in unclassified iron-sulfur proteins. The sliding window HMMs were constructed from 60-residue windows of the MSA built from 12 characterized RDase A proteins (Fig. S1). HMM windows overlapping the **PE**x_{13}**G**x_5**G** consensus retrieved 99% of class 1 sequences and 80% of class 2 sequences (Fig. 2, central, green and blue peak). Windows corresponding to positions of the **Y**x_{12}**F**x_3**LGY** and **G**xx**E**xx**R** consensus regions (residues 360 through 400) retrieved >95% of both class 1 and class 2 proteins (Fig. 2, right of center). By contrast, HMM windows overlapping the consensus sequences, **PE**x_{13}**G**x_5**G**, **Y**x_{12}**F**x_3**LGY**, **G**xx**E**xx**R**, retrieved less than 1% of the 985 unclassified iron-sulfur proteins. HMM windows overlapping the two iron-sulfur cluster-binding (ISB) motifs (residues 405 through 550) retrieved >90% of unclassified iron-sulfur proteins (Fig. 2, red peaks), thus providing evidence that the ISB motifs do not discriminate RDase superfamily sequences from unclassified iron-sulfur proteins. MSAs and sliding window RDase HMMs thus identified features unique to class 1 and class 2 sequences, delineating both classes as part of the superfamily and excluding unclassified iron-sulfur proteins. While specific residues or motifs involved in catalysis or cofactor binding cannot be determined without a

crystallographic analysis, the delineation of residues specific to the superfamily comprising RDases provide targets for functional analysis.

3.3.4 Amino acid sequence features unique to class 1.

Sliding window RDase HMMs delineated conserved amino acid sequence regions specific to class 1 sequences which were absent or sporadic among class 2 sequences and unclassified iron-sulfur proteins. Only class 1 proteins were retrieved by RDase HMM windows 1 through 225, corresponding to N-terminal Tat- or Sec-type signal peptides (Fig. 2, left on x-axis). HMM windows overlapping the conserved aromatic residues C-terminal of the two ISBs (residues 505 through 600) retrieved >95% of class 1 sequences. By contrast, HMM windows overlapping the aromatic residues C-terminal to the ISBs retrieved no more than 25% of class 2 sequences. Apparently, the C-terminal regions of class 1 sequences comprise a core structure encompassing the ISB clusters, and possibly mediating corrinoid co-factor binding. A predicted corrinoid-binding motif, **D**xx**H**xx**G**-x_n-**S**x**L**-x_n-**GG**, present in nine putative RDase A sequences (Hölscher et al., 2004; Futamata et al., 2009) coincides with the C-terminal core region. However, no regions of class 1 sequences matched HMMs from corrinoid-binding domains identified in methionine synthase and ribonucleotide reductase. Nonetheless, RDase HMM windows identified three likely structural domains in RDase A (class 1) sequences: an N-terminal signal-peptide region (~90 amino acids), a central region of unknown function (~60 amino acids), and a C-terminal region associated with the Fe-S clusters (~250 amino acids).

N-terminal sequence features characteristic of class 1 sequences were identified as either Sec or Tat signal peptide regions, where the latter type exhibited highly-conserved residues. An MSA of 229 class 1 sequences from complete or draft

genomes revealed a (S, N, or T)**RR**x**F**(M, L, or V)**K**xx**G** motif in the 80% Tat signal consensus, where residues at the first and sixth position exhibited similar physicochemical properties. After exclusion of sequences from the low-draft scaffolds available for Candidatus Dehalobium chlorocoercia strain DF-1, the motif **RR**x**F**x**K** appeared in the 90% Tat consensus. The six *Anaeromyxobacter* class 1 sequences, Adeh_0329, Adeh_0331, AnaeK_0341, AnaeK_0343, A2cp1_0353, and A2cp1_0355, and two of the three *Desulfomonile tiedjei* class 1 sequences, Desti_0783 and Desti_0785, did not align with the Tat motif consensus, yet exhibited Sec-type N-terminal regions under either Gram negative or Eukaryotic signal peptide models (www.cbs.dtu.dk/services/SignalP). The functional significance of the Sec-type N-terminal regions among RDases of *Anaeromyxobacter* and *D. tiedjei* is unclear; however, such divergence complicates efforts to specifically amplify RDase genes using a forward primer targeting the Tat consensus sequence (Krajmalnik-Brown et al., 2004).

3.3.5 B proteins associated with class 1 share conserved features.

RDase B proteins were unique to class 1 RDase A sequences, and exhibited a consistent set of structural features. Three transmembrane (TM) helices were detected in 190 RDase B sequences from genomes of the *Chloroflexi*, *Firmicutes*, *δ-Proteobacteria*, and *γ-Proteobacteria*. By comparison, seven RDase B sequences, one each from *Anaeromyxobacter dehalogenans* strains 2CP-C, 2CP-1, and K, *Desulfomonile tiedjei* strain DCB-1, *Sulfurospirillum* spp. and *Dehalococcoides mccartyi* strain BAV1 had two predicted TM helices. N-terminal Sec-type signal peptides were detected across all 197 three-helix and two-helix RDase B proteins using a eukaryotic signal peptide model (www.cbs.dtu.dk/services/SignalP). Based

upon the assumption that the N-terminal Sec signal is retained in the mature B proteins, 170 out of 190 three-helix RDase B proteins exhibit a membrane topology placing the N-termini periplasmic or extracellular (Fig. S2; model 1) while all seven two-helix RDase B proteins had cytoplasmic N-termini (Fig. S2; model 1). Alternatively, assuming the signal peptide is cleaved 180 of the three-helix B proteins had cytoplasmic N-termini (Fig. S2; model 2). The MSA of 197 predicted RDase B proteins revealed the 80% consensus motif (**W** or **Y**)$x_{2\text{-}4}$(**W** or **Y**), corresponding to the conserved **WY**x**W** motif of *Dehalococcoides mccartyi* RDase B proteins (Krajmalnik-Brown et al., 2004). The W/Y motif occurred within five residue positions of the start of TM helix regions in 176 out of the 197 B proteins. Moreover, in 170 intact B proteins (model 1) and 180 cleaved B proteins (model 2), the W/Y motif faced the cytoplasm. Hence, the conserved position and membrane orientation of the W/Y motif in B proteins suggests a critical, albeit unknown function.

3.3.6 Advantages of the HMM approach over Blast and automated gene annotation.

The RDase HMM and sequence curation approach used in this study outperformed Blast and RDase domain/structure profiles (e.g., PFAM) for identifying RDases and related sequences of the superfamily. Predicted RDases are typically annotated in RefSeq or GenBank fasta-formatted files as 'putative reductive dehalogenase', 'chloroethene reductive dehalogenase', or 'ortho-chlorophenol reductive dehalogenase'. Blast searches against GenBank (January, 2012) using *Dehalococcoides mccartyi* strain 195 TceA (DET00079) or the *Dehalobacter restrictus* strain PER-K23 PceA (CAD28790) as queries retrieved 477 and 425 proteins annotated as RDases, respectively (e-value < 0.00001). By contrast, the RDase HMM retrieved 544

GenBank proteins with RDase annotation. Moreover, the HMMs and the MSAs predict a larger set of superfamily sequences than indicated by protein sequence database annotation. For instance, 23 out of 476 class 1 sequences (excluding Candidatus Dehalobium chlorocoercia strain DF-1) lack RDase-related annotation, instead are annotated as '4Fe-4S ferredoxin' or 'hypothetical protein'. Among class 1 sequences with such incomplete annotation are the six putative CprA proteins of *Anaeromyxobacter dehalogenans*. RDases identified by TIGRFAM on RefSeq- and IMG-validated genomes (Selengut et al., 2007; Pruitt et al., 2007; Mavromatis et al., 2009) do not agree with counts determined by expert curation of *Dehalococcoides mccartyi* genomes (Seshadri et al., 2006; Kube et al., 2006; McMurdie et al., 2009). By comparison, the combined RDase HMM and sequence curation methods developed in this study match counts determined by expert curation of OHRB genomes (Table 1; Table S2). The combined RDase HMM and sequence curation approach developed in this study thus detects RDases with greater sensitivity than either Blast or TIGRFAM.

The RDase HMM and follow-up curation approach outperformed automated approaches in accurate detection and categorization of identified RDases. Established automated methods did not allow a clear inclusion threshold for proteins related to RDases. For instance, Blast searches using both PceA and TceA ranked non-RDases, such as ferredoxin hydrogenase, above class 1 sequences. NCBI TIGRFAM makes apparently spurious predictions of RDases on eukaryotic genomes, *Naegleria gruberi* and *Sus scrofa*, with e-values >6e-03. By comparison, the RDase HMM inclusion thresholds using e-values <2.5e-24 or bit scores >110 were validated by MSA analyses to retrieve only classes 1 and 2 protein sequences (lengths > 200 amino acids). Class 2 sequences exhibited inconsistent public database annotation where only 34 of the 59 class 2 sequences had annotation suggesting RDase function while the remaining 25

were annotated as “hypothetical protein”. The methodology of the current study explicitly assigned unknown or uncertain function to class 2 sequences while recognizing their evolutionary relatedness with membrane-associated class 1 (e.g., RDase A) sequences. In summary, the RDase HMM and follow-up analyses enabled more accurate and descriptive identification of proteins evolutionarily related to RDases than Blast, PFAM, and TIGRFAM.

3.3.7 Phylogenetic relationships within the RDase superfamily.

Phylogenetic analyses demonstrated sequence diversity within each class and provided further evidence for functional differentiation between the membrane-bound class 1 and cytoplasmic class 2 sequences. A bootstrapped phylogenetic tree containing 395 class 1, 55 class 2, and 9 unclassified sequences delineated relationships among distinct evolutionary lineages (e.g., clades) comprising sequences of the superfamily. Phylogenetic clusters with bootstrap support >40% (out of 200 replicates) were designated as clades (Fig. 3) and represented evolutionary lineages likely to share a common ancestor. Classes 1 and 2 sequences of the same clade or evolutionary lineage typically shared more than 30% pairwise amino acid sequence identity (Tables S6 and S7) while sequences compared across distinct class 1 lineages shared less than 30% pairwise global amino acid sequence identity. The analysis separated class 1 sequences into 11 clades (a_1 - k_1) plus three sequences lacking affiliation with any clade (Table S7), comprising 15 evolutionary lineages in total. Only the class 1 clades a_1, b_1, c_1, and j_1 contained biochemically characterized RDases while the PCE RDase (PrdA) of *Desulfitobacterium* sp. strain KBC1 did not cluster into a clade. Clade a_1 contained the characterized RDases of *Dehalococcoides mccartyi*, PceA, TceA, VcrA, BvcA, and CbrA, and comprised 152 sequences from genomes of organohalide-

respiring *Chloroflexi*, and 157 sequences from *Bacteria* lacking complete genome sequences. Class 2 sequences separated into six clades (a_2 - e_2) plus three sequences lacking affiliation with any clade (Table S8), 11 evolutionary lineages in total. Clade c_1 clustered with clade a_2 (support = 63% of 200 replicates) despite representing disparate protein lengths. While clade c_1 contains functionally characterized chlorophenol RDases (CprA), class 2 sequences of clade a_2 have unknown function and C-terminal NAD- and Fe-S-binding domains not identified by the RDase HMM (Fig. 1). At all other nodes shared between class 1 and class 2 lineages, bootstrap support for branching was below 40% or lacking entirely at the central node (Fig. 3). Despite low bootstrap support at the most ancestral nodes of the phylogeny, the nine unclassified iron-sulfur proteins clustered apart from all RDase superfamily evolutionary lineages. Thus, phylogenetic relatedness between classes 1 and 2 and distinctness from unclassified iron-sulfur proteins are apparent, yet it is not obvious whether class 1 or class 2 lineages are the most ancestral of RDase superfamily.

3.3.8 Evidence for convergent RDase evolution.

Comparisons between bootstrapped and constrained parsimony trees provided evidence that class 1 sequences (i.e., proteins with confirmed or predicted RDase functionality) originated multiple times (i.e., are polyphyletic) within the superfamily. Class 1 and class 2 clades shared common ancestral nodes in the bootstrapped trees, suggestive of polyphyly (Fig. 3 and Fig. 4). By comparison, constrained parsimony trees tested the null hypotheses of monophyly among proteins with similar predicted function (i.e., class 1 sequences) or identical characterized function (e.g., chlorinated ethene RDases). The first 226-sequence constrained tree tested the null hypothesis that all Tat-containing class 1 sequences are monophyletic (Fig. S3) and showed a

0.005 probability of representing the optimal phylogeny relative to the 226-sequence bootstrapped tree (Table 4, row 1). The second 226-sequence tree assumed that clade c_1 sequences originated membrane-associated function independently of other class 1 sequences (Fig. 4) and exhibited a 0.16 probability of representing the optimal phylogeny (Table 4, row 2). The low statistical support for the monophyly of Tat signal RDases suggests that RDase function in OHRB originated more than once within the superfamily.

Phylogenies of the 12 characterized RDases provided evidence that specific types of class 1 RDases, active towards similar substrates, have polyphyletic origins. The first 12-sequence constrained tree assumed a monophyletic origin of RDases with chloroaromatic substrates from clades a_1, c_1, and b_1 (Fig. S4) and exhibited a near zero probability relative to the bootstrapped 12-sequence tree (Table 4, row 3). A second constrained 12-sequence tree assumed a monophyletic origin of characterized PceA proteins (EC:1.97.1.8) from clades a_1, b_1, and j_1, similarly exhibiting a near-zero probability of representing the optimal tree (Table 4, row 4) The polyphyletic origin of RDase substrate affinity is further supported by the low global amino acid sequence identity shared between the CprAs of clades b_1 and c_1 (< 25% amino acid identity). In addition, characterized PceA proteins share no more than 30% global pairwise amino acid identity. These analyses indicate that RDases specific for similar organohalogen electron acceptors are represented in distinct evolutionary lineages. Conversely, RDases from the same clade may be active towards structurally distinct substrates, as is the case for *Desulfitobacterium* CprA5 and PceA proteins from clade b_1 (average 65% amino acid identities). Together, these analyses suggest that functionally characterized RDases are polyphyletic within class 1 of the superfamily and phylogenetic clustering may not accurately predict substrate range.

3.3.9 HGT of putative and characterized RDase genes.

The importance of HGT to the dissemination and evolution of known and putative RDases was investigated for the class 1 proteins. Inspection of the 283-sequence phylogeny (Fig. S5) and clade a_1 and b_1 trees (Figs. 5 and 6) served to predict lateral RDase gene dissemination. Evidence for HGT was based upon four criteria: (i) the RDase subtree representing the HGT event with > 50% bootstrap support at all nodes, (ii) RDase protein similarity greater than genomic housekeeping protein similarities, (iii) RDase-encoding genomic loci adjacent to predicted integrases or transposases, and (iv) topology discrepancies between the RDase subtree and the 16S rRNA gene phylogeny (Fig. S6). Applying these criteria, two subtrees, A and B, within the class 1 clade a_1 (Fig. 5) were predicted to represent HGT events crossing different genera of *Chloroflexi* OHRB. Protein sequences represented in subtree A, DehaBAV1_0284 from *Dehalococcoides mccartyi* strain BAV1 and DhcVS_96 and DhcVS_1336 from *Dehalococcoides mccartyi* strain VS, all shared an average 79% identity with Dehly_0275 from *Dehalogenimonas lykanthroporepellens* strain BL-DC-9. Protein sequences represented in subtree B from strain BAV1 (DehaBAV1_0112) and *Dehalococcoides mccartyi* strain CBDB1 (cbdb_A96) both shared 95% amino acid identity with ORF4 on the Candidatus Dehalobium chlorocoercia strain DF-1 scaffold 0604. By comparison, pairwise amino acid sequence identities for housekeeping proteins averaged 73% between *Dehalococcoides* and *Dehalogenimonas*, and 71% between *Dehalococcoides* and Candidatus Dehalobium chlorocoercia. The RDases of subtrees A and B thus exhibited higher similarities across the genera, *Dehalococcoides*, *Dehalogenimonas*, and Dehalobium than housekeeping proteins of their respective genomes. The topologies of subtrees A and B were congruent with the 16S rRNA gene tree topology (Fig. S6) and may provide evidence against HGT.

However, subtree B exhibits very short branch lengths (< 10% change in amino acid sites) compared with the 16S rRNA gene tree. Because lineage sorting can lead to agreement between the phylogenies of a species and a horizontally transferred gene (Than et al., 2007), tree topology agreement may not always rule out HGT events. Genomic association of subtrees A and B RDase gene loci with integrases or transposases (criterion iii) provided further evidence of HGT. On the strain BAV1 genome, a transposase *orfA-orfB* pair occurs upstream of the locus encoding the BAV1 sequence (DehaBAV1_0112) represented in subtree B. DhcVS_1336, from *D. mccartyi* strain VS, and Dehly_0275, from *Dehalogenimonas* strain BL-DC-9 (Fig. 5, subtree A), possess adjacent transposase *orfA-orfB* pairs, DhcVS_1332-DhcVS_1333 and Dehly_0278-Dehly_0279, respectively. Previous studies suggested that RDase exchange within the *Dehalococcoides* genus is common, and *vcrA* and *tceA*, both members of clade a_1 (Fig. 5), match highly similar homologs (>95% amino acid identity) across geographically-disparate environmental *Dehalococcoides* populations (Krajmalnik-Brown et al., 2006; McMurdie et al., 2011). In this study, subtrees A and B represent possible HGT events across distinct genera of the *Chloroflexi*, although the direction of transfer cannot be inferred.

A third subtree, Tr, in clade b_1 (Fig. 6), exhibited evidence of *pceA* HGT between *Desulfitobacterium* and *Dehalobacter*. The PceA proteins from *Dehalobacter restrictus* share 97% to 98% amino acid identity with the other known PceAs of clade b_1 while *Dehalobacter* housekeeping proteins (DNA gyrase, GroEL chaperonin, etc.) share 79% average amino acid identity with their homologs in *Desulfitobacterium*. Clade b_1 RDase gene HGT is supported by previous studies, where the *pceA* gene from *Desulfitobacterium* strain TCE1 was harbored on a transposon-like element (Maillard et al., 2005). Genes encoding clade b_1 PceA from

the *Geobacter lovleyi* genome exhibited guanine/cytosine content and codon usage atypical of the *Geobacteraceae*, suggesting horizontal transfer from an unidentified source (Wagner et al., 2012). While previous experimental studies demonstrated RDase transfer between strains of the same species (Krajmalnik-Brown et al., 2007; McMurdie et al., 2011), these findings together provide evidence that RDase genes are occasionally transferred across genera of OHRB.

3.4 Discussion.

3.4.1 Classes 1 and 2 belong to the same protein superfamily.

The application of HMM and MSA approaches demonstrated that over 450 protein sequences share similar features with experimentally characterized RDases, providing insights into shared RDases features. The central **PE**x_{13}**G**x_5**G** motif and C-terminal **Y**x_{12}**F**x_3**LGY** and **G**xx**E**xx**R** motifs shared across classes 1 and 2 satisfied established criteria for a protein superfamily (Bork and Koonin, 1996; Koonin et al., 1998) and excluded unclassified iron-sulfur proteins. The two domain-like regions shared among class 1 and class 2 sequences (Fig. 2) provided further evidence of membership in the same superfamily. Despite key similarities, RDase-related proteins could be divided into categories based upon presence/absence of N-terminal signal peptides and the associated RDase B protein. Class 1 sequences, including all characterized RDase A proteins, possess the N-terminal Tat- or Sec-type signal peptides enabling membrane export and associated RDase B proteins predicted to enable membrane association. By contrast, class 2 sequences lacked features mediating any known mechanism of interaction with the membrane. The accepted threshold for inference of shared function is 40% global pairwise amino acid identity (Tian and Skolnick, 2003). Here, the global amino acid identities between class 1 and class 2 sequences (counting length

differences) did not exceed 25%. As no member of class 2 has been biochemically characterized it is uncertain that class 2 proteins have RDase function. Yet, enzymes of the same superfamily often exhibit similar biochemical properties and only vary with respect to substrate specificity, metabolic role, or cellular localization (Park et al, 1998; Todd et al., 2001). Class 2 sequences likely do not share the same membrane-associated functional role as RDase A proteins; rather, class 2 sequences may exhibit RDase activity in the cytoplasm. For example, given that natural organohalides are abundant in seawater (Graedel and Keene, 1996; Gribble, 1998), class 2 proteins of marine *Bacteria* may serve in organohalide detoxification or aerobic degradation of halogenated aromatics as a carbon source. Future biochemical studies of class 2 proteins or growth studies of organisms which harbor class 2 proteins will be needed to elucidate class 2 function. Yet, the current computational study suggests cytoplasmic class 2 sequences may function as RDases in catabolism of chloroorganic compounds without coupling to respiration.

3.4.2 Conserved class 1 features may provide clues to RDase structure.

Motifs and domain-like regions conserved across class 1 serve as a basis for prediction of RDase superfamily structural or catalytic features. For instance, the conserved glycines or prolines in RDase superfamily motifs may correspond to type I or II beta-sheet hairpin turns (Bystroff and Parker, 1998; Chou, 2000) or terminating residues to alpha helices (Bystroff and Parker, 1998), and fulfill key structural functions. Variations within the ISB motifs may be related to differences in enzymatic properties or substrate specificity among RDases. The CprA of *Desulfitobacterium dehalogenans* was shown to bind a 3Fe-4S iron-sulfur cluster at the second ISB (van den Pas et al., 1999), contrasting with the 4Fe-4S of the second ISB in the

Dehalobacter restrictus PceA (Schumacher et al., 1997). Conserved residues thus provide evidence of common structural or catalytic features across the RDase superfamily, but exact function cannot be determined without RDase crystal structure information. The likelihood that RDases represent a superfamily of similar structures may in turn facilitate future efforts aimed at characterization RDase functionality and structure-function relationships.

3.4.3 RDase A localization and interactions with RDase B.

According to the current understanding of organohalide respiration, a B protein is a critical component to *in vivo* functionality of the catalytic RDase A (i.e., class 1). Translocation of the RDase A to the periplasm and membrane association via the associated RDase B appear to be essential requirements to coupling the reductive cleavage of carbon-halogen bonds to energy conservation and cell growth (Holliger et al., 1999; Smidt and de Vos, 2004). The *Sulfurospirillum multivorans* PceA and the *Dehalococcoides mccartyi* TceA were shown to be localized in the periplasm (John, Schmitz, et al., 2001; Nijenhuis and Zinder, 2005) yet, the role of the B subunit in anchoring or positioning of the RDase A has not been confirmed. In this study, W/Y motifs were found across 93% of RDase B proteins and were associated with the start of TM helices in 90% of RDase B, suggesting such aromatic-residue sites play a role in RDase A-RDase B interaction. W/Y motifs, particularly the **WY**x**W** sites of B proteins associated with clade a_1 sequences, resemble the protein-protein interaction domain motifs of SH3 domains or the WW domain (Ponting et al., 1999; Staub and Rotin, 1996). Yet, RDase B topology predictions placed the W/Y motifs on the cytoplasmic side of the membrane (Fig. S2), opposite the mature RDase A. A second 80% conserved site in the B proteins, a glutamate residue (E), occurs between the

second and third TM helices facing the periplasm and presumably able to interact with the A subunit. Glutamate residues have been characterized as a mechanism for protein-protein interactions through salt bridges, as is the case for the iron ABC transporter of *E. coli* (Braun and Hermann, 2007) Further experimental studies are needed to provide specific information about the interactions between RDase A, RDase B, and the cytoplasmic membrane.

3.4.4 The RDase HMM plus sequence curation provides improved genome annotation.

The current state of the art in genome annotation pipelines employs customized functional profiles while enabling manual curation of sequences with incomplete or non-descriptive annotation. While GenBank provides minimal validation of sequence annotation, RefSeq is built from validated gene-coding regions and annotated functions via PFAM and TIGRFAM (Benson et al., 2009; Pruitt et al., 2007). IMG at the Joint Genome Institute includes PFAM and TIGRFAM in its repertoire of tools for protein function prediction (Selengut et al., 2007; Finn et al., 2008; Mavromatis et al., 2009). Yet, none of these tools appear adequate to identify RDase superfamily proteins, many of which appear to be comprised of multiple domains. For example, class 1 proteins possess the N-terminal signal peptide region not included in the RDase models of PFAM and TIGRFAM. As the RDase HMM was built from full lengths of characterized RDases, it detected putative RDases more reliably than currently available RDase profiles. Given that OHRB often harbor multiple RDases (Table 1) the RDase HMM may be incorporated into genomics pipelines to screen OHRB (meta)genomes. Alternatively, the RDase HMM, together with the curation methods used in this study, could be employed in revisions of existing GenBank genome

records through IMG-ER (Markowitz et al., 2009). RDase profiles enabling accurate identification of RDase proteins and genes would in turn facilitate experimental studies on organohalide respiration or lead to enhanced models of RDase function prediction.

3.4.5 Implications of the polyphyletic origin of RDases.

Phylogenetic analyses provided evidence that characterized and putative class 1 RDases are polyphyletic, possibly as a result of convergent evolution. The higher statistical support for the bootstrapped 226-sequence phylogenies relative to the constrained parsimony phylogenies suggests RDase functionality may have evolved multiple times within a superfamily of related proteins. Likewise, 12-sequence phylogenies indicated multiple, independent origins of PCE RDases in clades a_1, b_1, and j_1, and chlorophenol RDases in clades b_1 and c_1. The sequence divergence within clade a_1, previously designated "cluster 1" (McMurdie et al., 2009), may reflect the structural diversity of organohalides degraded by *Chloroflexi*. Furthermore, the lack of detectable similarity between N-terminal Sec-type signals of *Anaeromyxobacter* class 1 sequences (clades h_1 and k_1) compared to class 1 sequences with Tat-type signals indicates multiple, independent origins of membrane export functions by the RDase superfamily. Such polyphyletic origins of distinct functional or structural domains have been shown for some superfamilies of multidomain enzyme systems such as transglutaminases and eukaryotic transcarboxylases (Makarova et al., 1999; Jordan et al., 2003). Given evidence of polyphyly for predicted and characterized RDase A, terms commonly used to describe RDase superfamily sequences, such as 'reductive dehalogenase homologous' (rdh) genes, may be accurate from the standpoint of shared structural features but should be used with caution. While protein

homology at the family level is easily detectable by pairwise identitites >40% (Dayhoff et al., 1975; Tian and Skolnick, 2003), describing homology or similarity between divergent members of a protein superfamily may require sophisticated tools (e.g., HMMs) to detect underlying shared structures (Dayhoff, 1976; Park et al, 1998). Hence, identification of homology among characterized or putative RDases may strongly depend upon the methodology for measuring similarity among RDases.

3.4.6 Organohalogen structural diversity as a factor in RDase evolution.

The structural diversity of naturally occurring organohalides may have contributed to the deeply branched, polyphyletic evolutionary history of RDases. At least 3,200 naturally occurring chloroorganic compounds have been identified (Gribble, 1998, 2003), all of which are potential electron acceptors for OHRB. The abundance in oceans and volcanic gasses of abiotically produced C-Cl compounds, such as chloromethanes (Graedel and Keene, 1996), indicate that organohalides have been a part of geochemical cycles since the origins of life. All characterized RDases catalyze the cleavage of the C-Cl bond associated with the reduction of the carbon and the release of H^+ and Cl^-. Yet, organohalogens exhibit enormous structural diversity, making prediction of RDase substrate specificity unlike function prediction for any other terminal reductase (e.g., nitrate reductases, cytochrome *c* oxidases, etc.). The apparent polyphyletic evolutionary history and the shear number of distinct RDases may reflect the great structural diversity of their organohalogen substrates (Gribble, 1998, 2003). Considering that many natural organohalides are not available for enrichment and isolation efforts, further computational studies combined with RDase enzymatic studies, and ultimately, an RDase crystallographic structure, may be essential to assigning function to many putative RDases.

3.4.7 Role of HGT in RDase gene dissemination.

While previous studies have provided evidence of intra-species transfer of RDase genes with specific functions (i.e., PceA, TceA) (Maillard et al., 2005; Krajmalnik-Brown et al., 2006), the current study identified instances of RDase HGT across genera. The tree comparative methodologies used in this study detected cross-genus HGT events indicating that HGT is not limited to strains of the same species or species of the same genus. Genomic comparisons provided supporting evidence that RDase-encoding mobile elements may occasionally cross genus boundaries of OHRB. Sequencing and curation of RDase gene-associated integrases, transposases, and repeats are commonly employed as a means to infer RDase HGT. For instance, identical flanking transposase genes appear to play a key role in genomic instability, and perhaps transfer, of PceA genes of clade b_1 (Maillard et al., 2005). Genes of the *vcrA* operon of *Dehalococcoides mccartyi* strains exhibit evidence for extensive horizontal dissemination, where genomic islands associated with the *ssrA*-gene (i.e., encoding transfer messenger RNA) may be involved in HGT (McMurdie et al., 2009; McMurdie et al., 2011). Similarly, *bvcA* in *Dehalococcoides mccartyi* strain BAV1 is flanked by duplicated tRNA genes (McMurdie et al., 2009). Moreover, the occurrence of highly similar *tceA* genes (>95% nucleotide identity) across geographically dispersed locations across North America and Europe suggest recent transfer and dispersal (Krajmalnik-Brown et al., 2006). In this study, the transposase *orfA*-*orfB* pairs adjacent to RDase genes involved in the two predicted cross-genus HGT events in clade a_1 thus represent one of possibly many genomic elements involved in RDase dissemination. Other, yet to be characterized OHRB genomic elements may play a role in the distribution of RDase superfamily sequences, and play roles in the adaptation of OHRB to anthropogenic contaminants. Despite the environmental

distribution of highly similar RDases, the detection of only three RDase HGT events crossing genera suggests that RDase superfamily sequences are rarely exchanged outside related strains of the same species.

3.4.8 Shortcomings of the tree comparative methodologies in detecting RDase HGT.

The tree comparisons employed in this study may not have provided an optimal method for detection of RDase HGT among related microbial genera, leading to possible underestimates in incidence of RDase gene HGT. Apparent vertical inheritance of RDase genes may have masked actual RDase gene HGT due to lineage sorting among the large numbers duplicated RDase genes on *Chloroflexi* genomes (Than et al., 2007). While cross-phyla or cross-domain HGT are readily detected by phylogenetic analyses, horizontal transfer of genes within a genus or between related genera can mimic vertical inheritance (Bieko et al., 2005; Andam and Gogarten, 2011). In this study, detection of three cross-genera and no cross-phyla HGT events on a phylogeny of nearly 300 RDases demonstrates HGT biased towards evolutionarily related strains (Andam et al., 2010; Andam and Gogarten, 2011). Taxonomic distance appears to comprise a strong barrier to RDase gene transfer, indicating RDase gene HGT is "biased" in favor of closely related species and strains (Andam et al., 2010), suggesting a biological basis for the low apparent incidence of RDase gene HGT. The respiratory functionality of RDases depends on a number of other genes, encoding functions for acquisition of the corrinoid cofactor, protein folding and assembly, etc., thus restricting RDase gene transfers to organisms with similar gene repertoire.

3.5 Conclusions and implications.

The HMM/MSA and sequence curation approaches employed in the current study advance insights into RDase structure, function, evolution, and improve genome annotation methodology. RDase A proteins (class 1) appear to be comprised of three domains: an N-terminal signal-peptide region (~90 amino acids), a central region of unknown function (~60 amino acids), and a C-terminal region associated with the Fe-S clusters (~250 amino acids). Class 2 proteins lack either the Tat- or Sec-type N-terminal region, yet share the central and C-terminal regions with class 1 and satisfy established criteria for inclusion in the same superfamily. Due to the likely multi-domain structure of RDase superfamily proteins, the combined HMM and sequence curation approach in this study provided more accurate identification of RDases than available through public genome annotation. Phylogenetic analyses conducted in this study inferred that membrane associated functionality of the class 1 proteins (RDase A) originated multiple times independently (i.e., polyphyletic origins). In addition, RDase substrate specificities for PCE and chlorophenols exhibit polyphyletic origins. RDases are also laterally transferred via mobile elements, though RDase HGT appears to be restricted to related species or genera (biased HGT). RDase polyphyly and HGT together may play a role in OHRB adaptation to specific anthropogenic contaminants, such as PCE.

Computational descriptions of RDase genes in collaboration with future work promises to provide a more complete picture of organohalide respiration as a geochemical process and a tool for contaminant mitigation. Improved functional classification or prediction of RDases may guide future experiments employing molecular or metagenomic characterization of organohalide-respiration in the environment. Delineation of conserved RDase protein features could facilitate

determination of RDase structure. Finally, RDase evolutionary history may provide insights into organohalide cycling across the geological timescale, from the origins of organohalide respiration to recent adaptations of OHRB to anthropogenic contaminants.

3.6 Methods.

Out of 16 experimentally-characterized RDases (Table S1), 12 RDase amino acid sequences were aligned in muscle (Edgar, 2004) and converted to the HMM using a local installation of hmmer-3.0 (Durbin et al., 1998). Four of the characterized RDase sequences shared > 95% amino acid identity with other RDases and could be excluded from the MSA without affecting HMM sensitivity. Biochemically characterized RDases included in the MSA (Fig. S1) and used to build the HMM included the PCE reductases (PceA) from *Sulfurospirillum multivorans* (AAC60788) (Neumann et al., 2002) and *Dehalobacter restrictus* (CAD28790) (Maillard et al., 2003), three biochemically characterized chlorophenol reductases (CprA) from *Desulfitobacterium dehalogenans* (AAD44542) (van den Pas et al., 1999), *Desulfitobacterium chlororespirans* (AAL84925) (Krasotkina et al., 2001), and *Desulfitobacterium hafniense* strain PCP-1 (AAQ54585) (Thibodeau et al., 2004), plus PrdA (PCE-responsive) from *Desulfitobacterium* sp. strain KBC1 (Tsukagoshi et al., 2005). Also included were the PCE reductase PceA (AW40342) and the TCE reductase TceA (AAW39060) from *Dehalococcoides mccartyi* strain 195 (Magnuson et al., 1998; Magnuson et al., 2000), the vinyl chloride (VC) reductase VcrA (ACZ62391) and BvcA (ABQ17429) from strain VS (Müller et al., 2004) and strain BAV1 (Krajmalnik-Brown et al., 2004), respectively, the chlorobenzene reductase CbrA (CAI82345) from strain CBDB1 (Adrian et al., 2007), and the protein encoded

by a trichlorobenzene-responsive gene (cbdbA1624, CAI83644) (Wagner et al., 2009) from strain CBDB1. The resulting RDase HMM was then queried against GenBank, InterPro, and SwissProt at http://mobyle.pasteur.fr and against genome project sequences from Candidatus Dehalobium chlorocoercia strain DF-1 and the *Dehalobacter* CF-50 metagenome (accession number pending).

A local, curated database of proteins was built based upon sequences retrieved by the RDase HMM with bit scores > 32 (or e-value < 0.0001). Here, bit scores indicate log-odds similarity derived from the sum of matches between HMM residue positions and sequences retrieved by the HMM (Eddy, 1998). To accurately categorize sequences based upon conserved N-terminal or C-terminal amino acid features, protein sequences <200 amino acids in length were removed from the set initially retrieved by the RDase HMM. An additional HMM was built using the N-terminal TAT domain (pfam10518) seed alignment (janelia.pfam.org) (Finn et al., 2010), and searched against the local, curated database. For sequences lacking a match with the TAT HMM, the TATp (www.cbs.dtu.dk/services/TatP/) and SignalP (www.cbs.dtu.dk/services/SignalP) servers were used to validate presence or absence of N-terminal membrane export signals. To search for cobalamin-binding domains, HMMs were constructed from seed alignments of the methionine synthase domains (pfam02607 and pfam02310) and ribonucleotide reductase (pfam02867). B proteins were inferred by length (80 to 100 amino acids), detection of transmembrane (TM) helices (www.cbs.dtu.dk/services/TMHMM), and detection of N-terminal membrane export signals. 197 predicted B proteins were aligned in muscle to visualize the W/Y motif **WY**x**W** (Hölscher et al., 2004; Krajmalnik-Brown et al., 2004; Edgar, 2004). Positions of W/Y motifs relative to TM helices were determined using custom Perl scripts. Two additional HMMs were constructed as models of non-RDase iron-sulfur

proteins based upon MSAs of archaeal polyferredoxins (Table S5, A) and iron-sulfur proteins encoded on genome of organisms known to lack organohalide respiration activity (Table S5, B). Windowed RDase HMMs were built by splitting the 12-sequence RDase MSA into 60-residue windows with 5-residue offsets using a custom Perl script. The resulting 120 HMM windows were then queried against the local, curated database. Domains and amino acid motifs shared among RDases were characterized using multiple sequence alignments (MSAs) built in Muscle (Edgar, 2004). To identify amino acid motifs specific to class 1 or class 2 of the RDase superfamily, MSAs were edited to contain sequences of only a single class. Conserved regions in the MSAs were identified in Gblocks (Castresana, 2000) and consensus sequences were calculated and visualized using showalign (Rice et al., 2000). Global pairwise amino acid identities were determined using EMBOSS needleall (Rice et al., 2000) while local pairwise amino acid identities were determined using Blast (Altschul et al., 1990).

To infer phylogenies over the set of sequences in the four MSAs, the tree-building packages Phylip, PAUP, and TreePuzzle were employed. Bootstrapped replicates of 100 to 200 were made for the 459-sequence, 283-sequence, and 226-sequence, and 12-sequence MSAs to infer topology by neighbor-joining (nj) consensus in Phylip (Felsenstein, 1989). Trees were plotted and visualized using the Interactive Tree of Life (iTOL) tool (Letunic and Bork, 2007). To test statistical likelihood of monophyly among evolutionary lineages, PAUP was used to construct constrained parsimony trees from the 226-sequence alignment and from the alignment of the 12 characterized RDases (Swofford, 2002). Constrained trees representing the null hypothesis of RDase monophyly, were compared with Phylip bootstrapped trees,

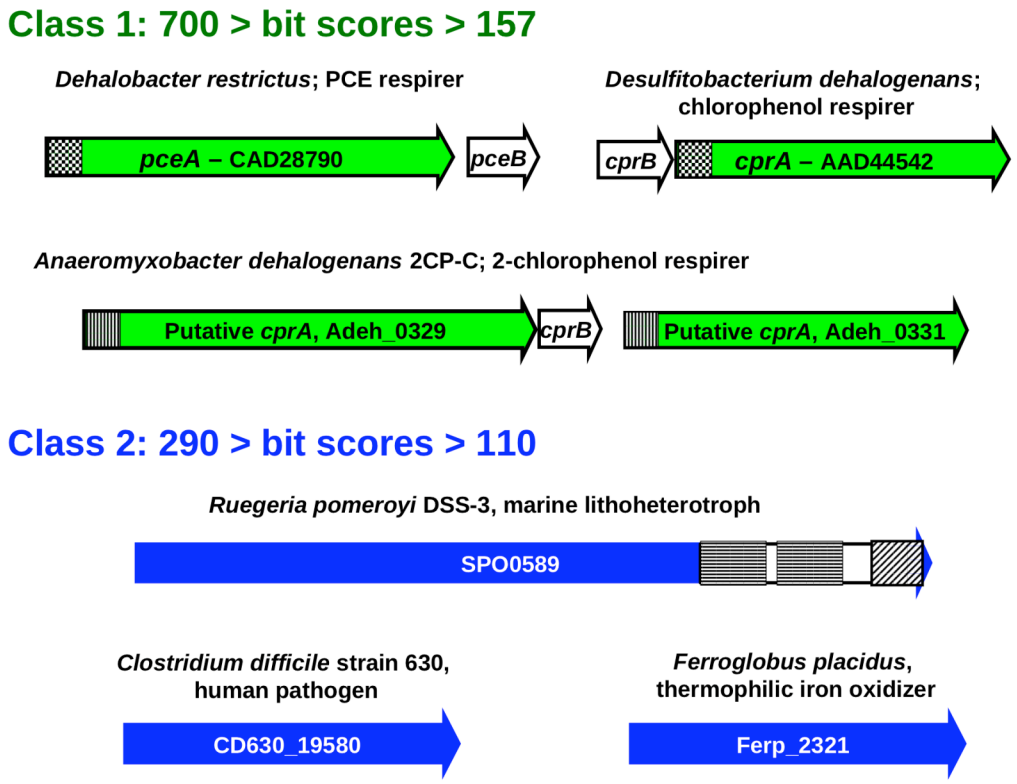
representing the test hypothesis of RDase polyphyly, using TreePuzzle as described (Schmidt and Haeseler, 2003; Jordan et al., 2003).

To infer instances of RDase HGT across the genus or phylum levels, four quantitative criteria were applied to the RDase protein phylogenetic trees to identify possible horizontally transferred pairs of class 1 RDase superfamily sequences. First, the RDase phylogeny subtree representing the potential HGT event was required to have > 50% bootstrap support at internal nodes and the most recent shared ancestral node. Second, amino acid identities between sequences represented in the subtree were required to be higher than the average amino acid identities of the nine housekeeping proteins DnaA, GyrA, GyrB, RpoB, EF-G, EF-Tu GroEL, GltB, and DnaK encoded on the respective genomes. Third, RDase-encoding loci represented in the tree were required to exhibit evidence of association with a mobile genomic element (i.e., adjacent integrases or transposases). Fourth, there was to be topology disagreement between the RDase subtree and the 16S rRNA gene tree of the RDase host genomes. Bootstrapped replicates of 1,000 were made from an alignment of 27 16S rRNA genes representing 14 OHRB genomes and a non-OHRB Archaeon, in Phylip (Felsenstein, 1989).

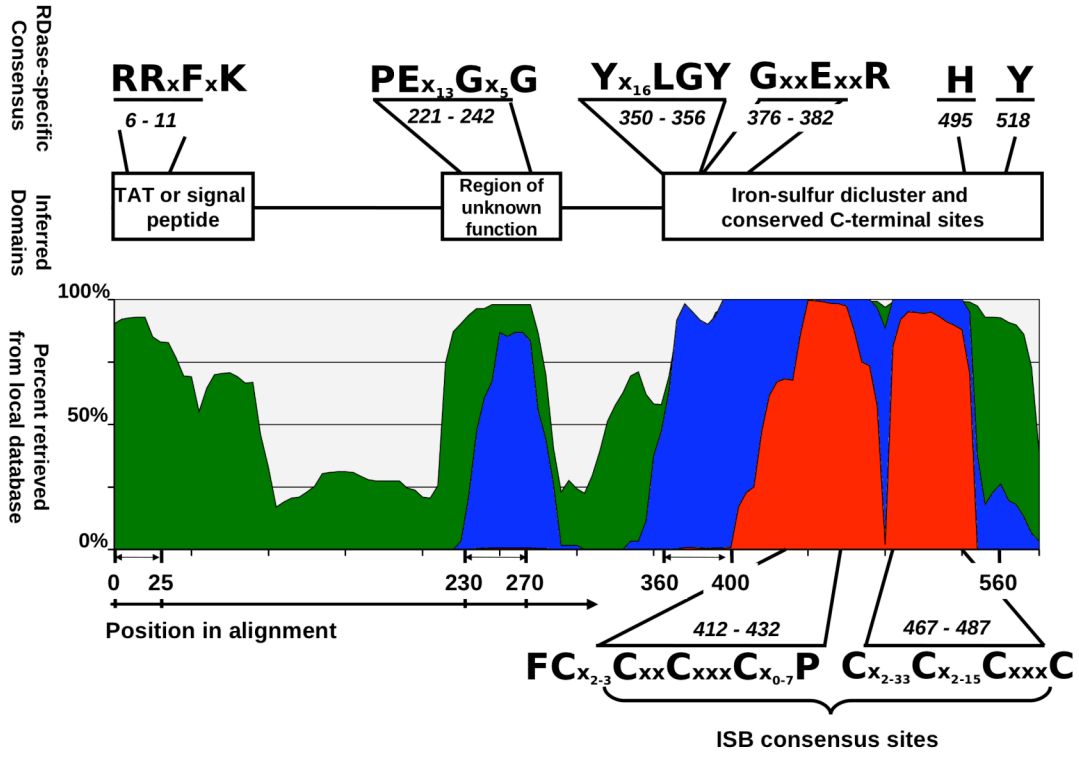
3.7 Chapter 3 Acknowledgements.

This research was supported by the Strategic Environmental Research and Development Program (SERDP) (project ER-1586).

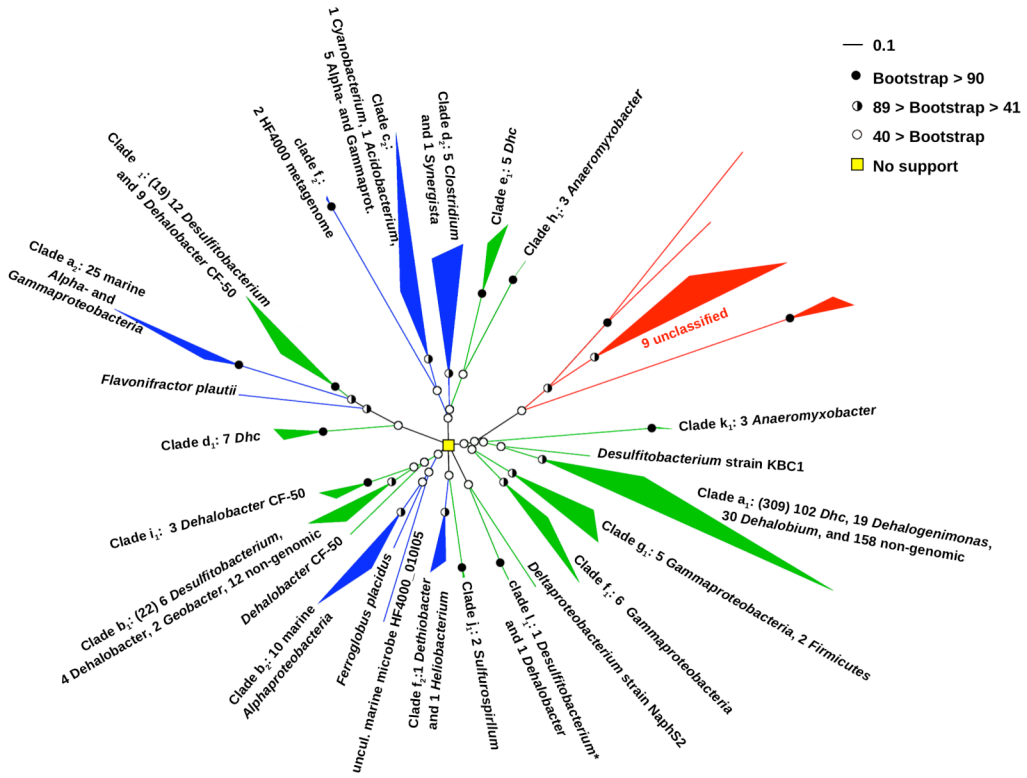
3.8 Chapter 3 Figures.



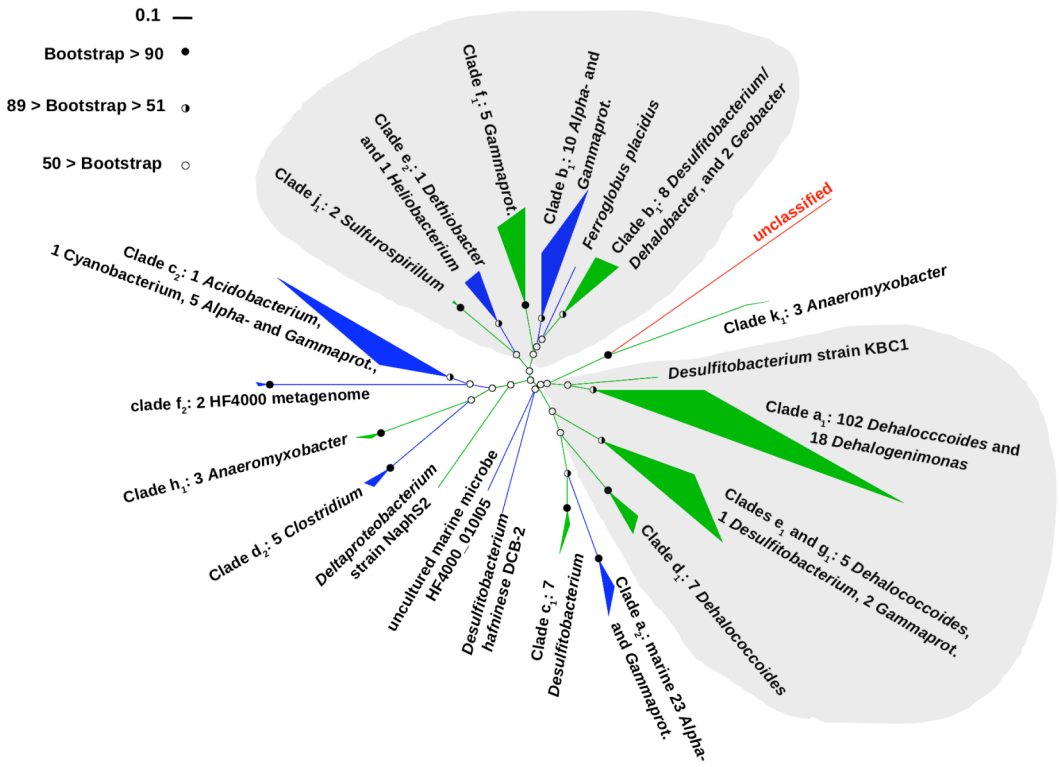
3.8.1 Figure 1. Features of representative class 1 and class 2 sequences retrieved by the RDase HMM. Class 1 sequences (green) exhibited distinct structural and genomic features from class 2 sequences (blue). Distinguishing features of class 1 sequences identified by sequence curation steps apart from the RDase HMM (black-white) include the Tat-type N-terminal signal peptide (checkered fill) and the Sec-type N-terminal signal peptide (vertical bars). Class 1 sequences (i.e., characterized and putative RDase A proteins) were encoded at loci associated with a gene encoding a predicted RDase B membrane-bound protein (white). Among a group of class 2 sequences from non-organohalide-respiring, marine α- and γ-*Proteobacteria*, C-terminal domain regions were identified separately from the RDase HMM, two flavin-binding domains (pfam00970 and pfam00175, horizontal bars) and a 2Fe-2S domain (pfam00111, diagonal bars).



3.8.2 Figure 2. Core structural regions of RDase superfamily proteins inferred by windowed RDase HMMs. The y-axis indicates percentages of sequences retrieved from among class 1 (green) and class 2 (blue) by windowed RDase HMMs where peaks in percentages correspond with motifs in the 90% consensus of the 283-sequence MSA. Positions of RDase HMMs are shown on the x-axis as corresponding residue positions from the MSA of the 12 characterized RDases. Italicized numbers below motifs indicate position in the *Dehalobacter restrictus* strain PER-K23 PceA. Unclassified sequences (red) were retrieved by the windowed RDase HMM only at positions corresponding to the iron-sulfur motifs.

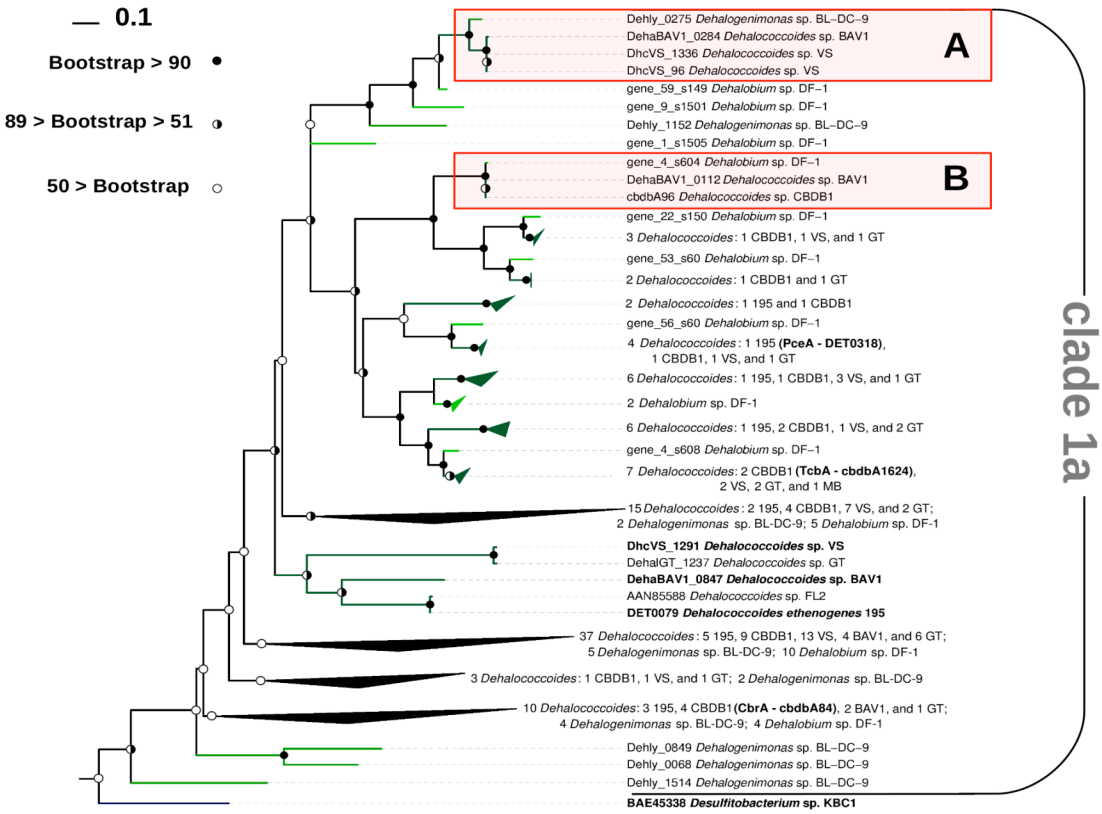


3.8.3 Figure 3. Evolutionary lineages (clades) of the RDase superfamily determined by bootstrapped phylogenetic reconstruction from 459 sequences aligned at 126 residues. Bootstrapped neighbor-joining tree based upon a MSA of 395 class 1 sequences (green), 55 class 2 sequences (blue) and nine unclassified iron-sulfur proteins (red). Only class 1 clades a_1, b_1, c_1, and j_1, have biochemically characterized representatives. *Dhc* denotes sequences from the genomes of *Dehalococcoides mccartyi.* The central, boxed node represents unresolved topology. Bootstrap values were computed as percentages of 200 replicates of the MSA and the scale bar in the upper right corner corresponds to amino acid changes at 10% of aligned sites.

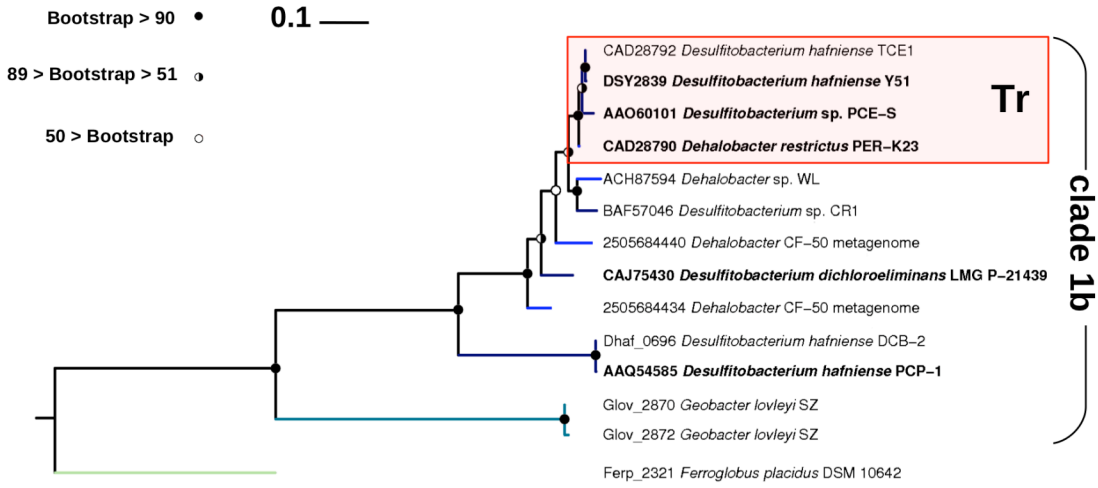


3.8.4 Figure 4. Bootstrapped tree of 226 sequences aligned at 159 residues.

Optimal topology of bootstrapped tree is shown with assumed monophyletic clades for hypothesis 2 (Table 4, row 2) shaded in gray. The bootstrapped tree has likelihood score = – 32358 while the hypothesis 2 tree, which also shows class 2 polyphyly, has likelihood score = – 32400. The scale bar in the upper left hand corner corresponds to amino acid changes at 10% of aligned sites.



3.8.5 Figure 5. RDase HGT within clade a_1 (*Chloroflexi*). The highlighted subtrees A and B represent possible HGT events between *Dehalococcoides mccartyi* strains BAV1, VS, and *Dehalogenimonas lykanthroporepellens* strain BL-DC-9 or *Dehalococcoides mccartyi* strains BAV1, CBDB1, and *Dehalobium chlorocoercia* strain DF-1, respectively. Biochemically characterized RDases, all from *Dehalococcoides mccartyi*, are shown in boldface. Bootstrap values are represented by open circles (support < 49%), half-closed circles (50% < support < 89%), and black circles (90% < support). The scale bar in the upper right corner corresponds to amino acid changes at 10% of aligned sites.



3.8.6 Figure 6. RDase HGT within clade b_1 (*Firmicutes*). Subtree Tr represents a group of possible transposon-mediated HGT events involving biochemically characterized RDases (boldface font). Bootstrap values are represented by open circles (support < 49%), half-closed circles (50% < support < 89%), and black circles (90% < support). The scale bar in the upper right corner corresponds to amino acid changes at 10% of aligned sites.

3.9 Chapter 3 Tables.

3.9.1 Table 1: RDase superfamily proteins of OHRB genomes. RDases identified by RDase HMM on OHRB (meta)genomes. For comparison, genomic RDases identified by the RDH TIGRFAM profile, TIGR02486, are shown for NCBI and IMG. Class 2 sequences (bold type) exhibit differences in sequence lengths compared to the average length of 489 residues for class 1 sequences. (N/A = Not available)

Organism	Genome Accession	Retrieved by RDase HMM and validated by subsequent analyses			RDH TIGRFAM profile	
		Sequence count	Avg. bit score	Lengths (aa)	NCBI	IMG
Dehalococcoides mccartyi str. 195	CP000027	17 class 1	388	455 - 554	18	19
Dehalococcoides mccartyi str. CBDB1	AJ965256	32 class 1	372	455 - 532	32	33
Dehalococcoides mccartyi str. BAV1	CP000688	11 class 1	342	455 - 532	12	11
Dehalococcoides mccartyi str. VS	CP001827	36 class 1	362	429 - 532	37	36
Dehalococcoides mccartyi str. GT	CP001924	20 class 1	405	455 - 519	20	20
Dehalogenimonas str. BL-DC-9	CP002084	19 class 1	331	340 - 493	18	21
Dehalobium str. DF-1	Pending	30 class 1	354	416 - 544	N/A	
Desulfitobacterium hafniense str. Y51	AP008230	1 class 1	601	551	1	1
Desulfitobacterium hafninese str. DCB-2	CP001336	6 class 1 **1 class 2**	377 **289**	345 – 550 **352**	7	7
Desulfitobacterium sp. str. PCE-S	N/A	1 class 1	594	551	N/A	
Desulfitobacterium sp. str. PCP-1	N/A	1 class 1	585	547	N/A	
Desulfitobacterium dichloroeliminans LMG P-21439	AGJE01-000036	1 class 1	561	551	N/A	
Desulfitobacterium dehalogenans ATCC 51507	CP003348	4 class 1 **2 class 2**	394 **262**	447 - 468 **403 - 690**	N/A	
Desulfitobacterium chlororespirans	N/A	2 class 1	429	450 - 456	N/A	
Desulfitobacterium sp. str. KBC1	N/A	2 class 1	481	447 - 463	N/A	
Desulfitobacterium sp. str. PCE1	N/A	1 class 1	456	447	N/A	
Dehalobacter restrictus str. PER-K23	N/A	1 class 1	606	551	N/A	
Dehalobacter metagenome CF-50	Pending	17 class 1	351	429 - 551	N/A	
Geobacter lovleyi str. SZ	CP001089	2 class 1	315	Both 514	2	2
Anaeromyxobacter dehalogenans str. 2CP-C	CP000251	2 class 1	157	484 - 640	2	1
Anaeromyxobacter dehalogenans str. 2CP-1	CP001359	2 class 1	157	474 - 640	2	0
Anaeromyxobacter dehalogenans str. K	CP001131	2 class 1	156	484 - 640	2	0
Desulfomonile tiedjei str. DCB-1	CP003360	3 class 1	230	351 - 559		
Sulfurospirillum multivorans	N/A	1 class 1	493	501	N/A	
Sulfurospirillum halorespirans	N/A	1 class 1	493	501	N/A	

3.9.2 Table 2. RDase superfamily proteins from organisms lacking characterization as OHRB. Representative class 1 and class 2 sequences were retrieved by RDase HMM from genomes of strains not known for organohalide-respiration. Genomic RDases identified by the RDH TIGRFAM profile, TIGR02486, are shown for NCBI (www.ncbi.nlm.nih.gov) and IMG (www.img.jgi.gov).

Organism	Genome Accession	Retrieved by RDase HMM and validated by subsequent analyses			Assigned RDH TIGRFAM profile	
		Sequence count	Average bit score	Lengths (aa)	NCBI	IMG
Shewanella sediminis str. HAW-EB3	CP000821	5 class 1	242	468 - 507	5	5
Vibrio scophthalmi str. LMG19158	AFWE00000000	2 class 1	230	474 - 516		N/A
Vibrio sp. str. RC586	ADBD00000000	1 class 1	250	470	0	1
Vibrio sp. str. N418	AFWD00000000	1 class 1	185	504		N/A
Vibrio ichthyoenteri ATCC700023	AFWF00000000	1 class 1	189	504		N/A
Photobacterium profundum str. 3TCK	AAPH00000000	1 class 1	197	506	0	1
Gammaproteobacterium str. NOR51-B	ACCY00000000	4 class 2	192	328 - 449	0	2
Ruegeria pomeroyi str. DSS-3	CP000031	3 class 2	188	328 - 1090	1	2
Rhodobacterales str. HTCC2083	ABXE000000000	2 class 2	232	390 - 1068	0	2
Clostridium difficile str. 630	AM180355	1 class 2	157	345	0	0
Dethiobacter alkaliphilus str. AHT 1	ACJM00000000	1 class 2	287	407	0	1
Heliobacterium modesticaldum str. Ice1	CP000930	1 class 2	284	405	1	1
Ferroglobus placidus DSM 10642	CP001899	1 class 2	270	406	1	1

N/A = not available

3.9.3 Table 3. Conserved RDase motifs. Amino acid sites exhibiting at least 90% conservation across class 1 sequences (membrane associated), class 2 sequences (cytoplasmic), and unclassified sequences.

	Domain-like regions		
	N-terminal	**Central**	**C-terminal**
Class 1 characterized	**RR**x**F**	**PE**x_{13}**G**x_5**G**	**Y**x_{12}**F**x_3**LGY**, **G**x**GE**xx**R**, **ISB1**, **ISB2**, **H**, **Y**
Class 1 putative	**RR**x**F**	**PE**x_{13}**G**x_5**G**	**Y**x_{12}**F**x_3**LGY**, **G**xx**E**xx**R**, **ISB1**, **ISB2**, **H**, **Y**
Class 2		**PE**x_{13}**G**x_5**G**	**Y**x_{13}**IR**x_2**GY**, **G**xx**E**xx**R**, **ISB1**, **ISB2** *
unclassified			**ISBs** only, numbers variable

* clade a_2 sequences exhibit additional domains C-terminal to the two ISBs, the flavin-binding domains, pfam00970 and pfam00175, and a 2Fe-2S domain, pfam00111.

3.9.4 Table 4. Scoring of constrained trees, representing null hypotheses of RDase A monophyly. One-sided Kishino-Hasegawa statistic represents the probability that the constrained tree represents the true phylogeny relative to bootstrapped neighbor-joining trees. The 226-sequence – Null hypothesis 2 (row 2) assumes limited polyphyly in which respiratory class 1 sequences originated twice. Log-likelihood of the 226-sequence bootstrapped tree, representing the test hypothesis of polyphyletic class 1 sequences, is -32358.25. Log-likelihood of the 12-sequence bootstrapped tree, representing the test hypothesis of polyphyletic substrate ranges, is -13379.58.

	Tree tested	Evolutionary scenario tested	Log-likelihood of tree.	One-sided Kishino-Hasegawa	Reject null hypothesis at 0.05?
1	**226-sequence – Null hypothesis 1:**	All class 1 sequences with the TAT-motif are monophyletic.	-32484.62	0.005	Yes
2	**226-sequence – Null hypothesis 2:**	Clade c_1 originated respiratory functionality independently of other class 1 sequences. Clades b_2 and f_2 share ancestry with clades a_1, b_1, and d_1-j_1.	-32400.43	0.164	No
3	**12-sequence – Null hypothesis 1:**	Chloroaromatic-substrate RDases: CprAs, CbrA, and cbdbA1624 are assumed monophyletic.	-13953.99	0.0005 >	Yes
4	**12-sequence – Null hypothesis 2:**	PceAs of *Dehalobacter restrictus*, *Sulfurospirillum multivorans*, *Dehalococcoides mccartyi* strain 195, and PrdA of *Desulfitobacterium* sp. strain KBC1 assumed monophyletic	-13925.07	0.0005 >	Yes

3.10 Chapter 3 References.

Adrian, L., Rahenführer, J., Gobom, J., and Hölscher, T. (2007) Identification of a Chlorobenzene reductive dehalogenase in *Dehalococcoides* sp. Strain CBDB1. *Applied and Environmental Microbiology* **73**: 7717-7724.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Andam, C. P. and Gogarten, J. P. (2011) Biased gene transfer in microbial evolution. *Nature Reviews Microbiology* **12**:543-555

Andam, C. P., Williams, D., and Gogarten, P. J. (2010) Biased gene transfer mimics patterns created through shared ancestry. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 10697-10684.

Banerjee, R., and Ragsdale, S.W. (2003) The many faces of vitamin B_{12}: Catalysis by cobalamin-dependent enzymes. *Annual Reviews of Biochemistry* **72**: 209-247.

Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005) Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**:14332-14337.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2008) GenBank. *Nucleic Acids Research* **37**: 26-31.

Berks, B.C., Palmer, T., and Sargent, F. (2005) Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology* **8**: 174-181.

Bork, P., and Koonin, E. V. (1996) Protein sequence motifs. *Current Opinion in Structural Biology* **6**: 366-376.

Braun, V. and Herrmann C. (2007). Docking of the Periplasmic FecB Binding Protein to the FecCD Transmembrane Proteins in the Ferric Citrate Transport System of *Escherichia coli*. *Journal of Bacteriology* **189:** 6913–6918.

Bystroff, C., and Parker, D. (1998) Prediction of local structure in proteins using a library of sequence-structure motifs. *Journal of Molecular Biology* **281**: 565-577.

Castresana, J. (2000) Selection of conserved blocks from multiple sequence alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540-552.

Chou, K.-C. (2000) Prediction of tight turns and their types in proteins. *Analytical Biochemistry* **286**: 1-16.

Christiansen, N., Ahring, B.K., Wohlfarth, G., and Diekert, G. (1998) Purification and characterization of the 3-chloro-4-hydroxy-phenylacetate reductive dehalogenase of *Desulfitobacterium hafniense*. *FEBS Letters* **436**: 156-162.

Cole, J.R., Cascarelli, A.L., Mohn, W.W., and Tiedje, J.M. (1994) Isolation and characterization of a novel bacterium growing via reductive dehalogenatioin of 2-chlorophenol. *Applied and Environmental Microbiology* **60**: 3536-3542.

Dayhoff, M. O. (1976) The origin and evolution of protein superfamilies. *Federation Proceedings* **35**: 2132-2138.

Dayhoff, M. O., McLaughlin, P. J., Barker, W. C., Hunt, L. T. (1975) Evolution of sequences within protein superfamilies. *Naturwissenschaften* **62**: 154-161.

Duhamel, M., Mo, K., and Edwards, E.A. (2004) Characterization of a highly enriched *Dehalococcoides*-containing culture that grows on vinyl chloride and trichloroethene. *Applied and Environmental Microbiology* **70**: 5538-5545.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998) Biological sequence analysis: Probabalistic models of proteins and nucleic acids. *Cambridge University Press*.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics Review* **14**: 755-763.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Felsenstien, J. (1989) PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics* **5**: 164-166.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E. et al. (2010) The Pfam protein families database. *Nucleic Acids Research* **38**: 211-222.

Furukawa, K., Suyama, A., Tsuboi, Y., Futagami, T., and Goto, M. (2005) Biochemical and molecular characterization of a tetrachloroethene dechlorinating *Desulfitobacterium* sp. strain Y51: a review. *Journal of Industrial Microbiological Technology* **32**: 534-541.

Futamata, H., Kaiya, S., Sugawara, M., and Hirashi, A. (2009) Phylogenetic and transcriptional analysis of a tetrachloroethene-dechlorinating "*Dehalococcoides*" enrichment culture TUT2264 and its reductive-dehalogenase genes. *Microbes and Environments* **24**: 330-337.

Graedel, T. and Keene, W.C. (1996) The budget and cycle of Earth‘s natural chlorine. *Pure and Applied Chemistry*. **68:** 1689-1697.

Gribble, G.W. (1998) Naturally Occurring Organohalogen Compounds. *Accounts of Chemical Research* **31**: 141-152.

Gribble, G.W. (2003) The diversity of naturally produced organohalogens. *Chemosphere* **52**: 289-297.

Grostern, A., Duhamel, M., Dworatzek, S., and Edwards, E.A. (2010) Chloroform respiration to dichloromethane by a *Dehalobacter* population. *Environmental Microbiology* **12**: 1053-1060.

Häggblom, M.M., Ahn, Y.-B., Fennell, D.E., Kerkhof, L.J., and Rhee, S.-K. (2003) Anaerobic dehalogenation of organohalide contaminants in the marine environment. *Advances in Applied Microbiology* **53**: 61-84.

He, J., Ritalahti, K.M., Yang, K.-L., Koenigsberg, S.S., and Löffler, F.E. (2003) Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* **424**: 62-65.

Holliger, C., Hahn, D., Harmsen, H., Ludwig, W., Schumacher, W., Tindall, B., Vasquez, F., Weiss, N., and Zehnder, A. J. B. (1998) *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetra- and trichloroethene in an anaerobic respiration. *Archives of Microbiology* **169**: 313-321.

Holliger, C., Schraa, G., Stams, A.J.M., and Zehnder, A.J.B. (1993) A highly purified enrichment culture couples the reductive dechlorination of tetrachloroethene to growth. *Applied and Environmental Microbiology* **59**: 2991-2997.

Hölscher, T., Krajmalnik-Brown, R., Ritalahti, K.M., Wintzingerode, F.v., Görisch, H., Löffler, F.E., and Adrian, L. (2004) Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*. *Applied and Environmental Microbiology* **70**: 5290-5297.

John, M., Schmitz, R.P.H., Westermann, M., Richter, W., Diekert, G. (2006) Growth substrate dependent localization of tetrachloroethene reductive dehalogenase in *Sulfurospirillum multivorans*. Archives of Microbiology. **186:** 99-106.

Jordan, I.K., Henze, K., Federova, N.D., Koonin, E.V., and Galperin, M.Y. (2003) Phylogenomic analysis of the *Giardia intestinalis* transcarboxylase reveals multiple instances of domain fusion and fission in the evolution of biotin-dependent enzymes. *Journal of Molecular Microbiology and Biotechnology* **5**: 172-189.

Kim, S.-H., Harzman, C., Davis, J.K., Hutcheson, R., Broderick, J.B., Marsh, T.L., Tiedje, J.M. (2012) Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiology* **12:** 21.

Koonin, E. V., Tatusov, R. L., and Galperin, M. Y. (1998) Beyond complete genomes: from sequence to structure to function. *Current Opinion in Structural Biology* **8**: 355-363.

Krajmalnik-Brown, R., Sung, Y., Ritalahti, K.M., Saunders, F.M., and Löffler, F.E. (2007) Environmental distribution of the trichloroethene reductive dehalogenase gene (*tceA*) suggests lateral gene transfer among *Dehalococcoides*. *FEMS Microbiology Ecology* **59**: 206-214.

Krajmalnik-Brown, R., Hölsher, T., Thomson, I.N., Saunders, F.M., Ritalahti, K.M., and Löffler, F.E. (2004) Genetic identification of a putative vinyl chloride reductase in *Dehalococcoides* sp. Strain BAV1. *Applied and Environmental Microbiology* **70**: 6347-6351.

Krasotkina, J., Walters, T., Maruya, K.A., and Ragsdale, S.W. (2001) Characterization of the B_{12}- and iron-sulfur-containing reductive dehalogenase from *Desulfitobacterium chlororespirans*. *The Journal of Biological Chemistry* **276**: 40991-40997.

Kraütler, B., Fieber, W., Ostermann, S., Fasching, M., and Ongania, K.-H. (2003) The cofactor of tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans* is Norpseudo-B_{12}, a new type of a natural corrinoid. *Helvetica Chimica Acta* **86**: 3698-3716.

Kube, M., Beck, A., Zinder, S.H., Kuhl, H., Reinhardt, R., and Adrian, L. (2005) Genome sequence of the chlorinated compound-respiring bacterium *Dehalococcoides* species strain CBDB1. *Nature Biotechnology* **23**: 1269-1273.

Lee,P.K., Macbeth,T.W., Sorenson,K.S. Jr., Deeb,R.A. and Alvarez-Cohen,L. (2008) Quantifying genes and transcripts to assess the in situ physiology of '*Dehalococcoides*' spp. in a trichloroethene-contaminated groundwater site. *Applied and Environmental Microbiology* **74**: 2728-2739.

Letunic, I., and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127-128.

Löffler, F. E., Edwards, E. A. (2006) Harnessing microbial activities for environmental cleanup. *Current Opinion in Biotechnology* **17**: 274-284.

Löffler, F.E., Sanford, R.A., and Tiedje, J.M. (1996) Initial Characterization of a Reductive Dehalogenase from *Desulfitobacterium chlororespirans* Co23. *Applied and Environmental Microbiology* **62**: 3809-3813.

Ludwig, M. L., and Matthews, R. (1997) Structure-based perspectives on B_{12}-dependent enzymes. *Annual Review of Biochemistry* **66**: 269-313.

Magnuson, J.K., Romine, M.F., Burris, D.R., and Kingsley, M.T. (2000) Trichloroethene reductive dehalogenase from *Dehalococcoides ethenogenes*: Sequence of *tceA* and substrate range characterization. *Applied and Environmental Microbiology* **66**: 5141-5147.

Magnuson, J.K., Stern, R.V., Gosset, J.M., Zinder, S.H., and Burris, D.R. (1998) Reductive dechlorination of tetrachloroethene to ethane by a two-component enzyme pathway. *Applied and Environmental Microbiology* **64**: 1270-1275.

Maillard, J., Regeard, C., and Holliger, C. (2005) Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1. *Environmental Microbiology* **7**: 107-117.

Maillard, J., Schumacher, W., Vazquez, F., Regeard, C., Hagen, W.R., and Holliger, C. (2003) Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus. Applied and Environmental Microbiology* **69**: 4628-4638.

Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I-M.A., Chu, K., Kyrpides, N.C. (2009) IMG ER: a system for microbial annotation expert review and curation. *Bioinformatics* **25**: 2271-2278.

Makarova, K. S., Aravind, L., and Koonin, E. V. (1999) A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein Science* **8**: 1714-1719.

Mavromatis, K., Ivanova, N.N., Chen, I-M..A., Szeto, E., Markowitz, V.M., and Kypides, N.C. (2009) The DOE-JGI standard operating procedure for the annotations of microbial genomes. *Standards in Genomic Science* **1**: 63-67.

McMurdie, P.J., Hug, L., Edwards, E.A., Holmes, S., and Spormann, A.M. (2011) Site-specific mobilization of vinyl chloride respiration islands by a mechanism common in *Dehalococcoides*. *BMC Genomics* **12**: 287.

McMurdie, P.J., Behrens, S., Müller, J.A., Göke, J., Ritalahti, K.M., Wagner, R. et al. (2009) Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genetics* **5**: e1000714.

Miller, E., Wohlfarth, G., and Diekert, G. (1998) Purification and characterization of the tetrachloroethene reductive dehalogenase of strain PCE-S. *Archives of Microbiology* **169**: 497-402.

Mohn, W. W. and Kennedy, K. J. (1992) Reductive dehalogenation of chlorophenols by *Desulfomonile tiedjei* DCB-1. *Applied and Environmental Microbiology* **58**:1367-1370.

Müller, J.A., R., B.M., Abendroth, G.v., Meshulam-Simon, G., McCarty, P.L., and Spormann, A.M. (2004) Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. Strain VS and its environmental distribution. *Applied and Environmental Microbiology* **70**: 4880-4888.

Nakamura,K., Mizumoto,M., Ueno,T. and Ishida, H. (2006) Cloning and analysis of trichloroethene reductive dehalogenase gene and its detection by quantitative real-time PCR. *Environmental Engineering Research* **43**: 119-125.

Neumann, A., Scholz-Muramatsu, H., and Diekert, G. (1994) Tetrachloroethene metabolism of *Dehalospirillum multivorans*. *Archives of Microbiology* **162**: 295-301.

Neumann, A., Wohlfarth, G., and Diekert, G. (1996) Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*. *The Journal of Biological Chemistry* **271**: 16515-16519.

Neumann, A., Wohlfarth, G., and Diekert, G. (1998) Tetrachloroethene dehalogenase from *Dehalospirillum multivorans*: Cloning, sequencing of the encoding genes, and expression of the *pceA* gene in *Escherichia coli*. *Journal of Bacteriology* **180**: 4140-4145.

Neumann, A., Siebert, A., Trescher, T., Reinhardt, S., Wohlfarth, G., and Diekert, G. (2002) Tetrachloroethene reductive dehalogenase of *Dehalospirillum multivorans*: substrate specificity of the native enzyme and its corrinoid cofactor. *Archives of Microbiology* **177**: 420-426.

Ni, S., Fredrickson, J.K., and Xun, L. (1995) Purification and characterization of a novel 3-chlorobenzoate-reductive dehalogenase from the cytoplasmic membrane of *Desulfomonile tiedjei* DCB-1. *Journal of Bacteriology* **177**: 5135-5139.

Nijenhuis, I. and Zinder, S.H. (2005) Characterization of hydrogenase and reductive dehalogenase activities of *Dehalococcoides ethenogenes* Strain 195. Applied and Environmental Microbiology. **71:** 1664-1667.

Nishimura, M., Ebisawa, M., Sakihara, S., Kobayashi, A., Nakama, T., Okochi, M., and Yohda, M. (2008) Detection and identification of *Dehalococcoides* species responsible for *in situ* dechlorination of trichloroethene to ethene enhanced by hydrogen-releasing compounds. *Biotechnology and Applied Biochemistry* **51**: 1-7.

Nonaka, H., Keresztes, G., Shinoda, Y., Ikenaga, Y., Abe, M., Inatomi, K., Furukawa, K., Inui, M., and Yukawa, H. (2006) Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195. *Journal of Bacteriology* **188:** 2262-2274.

Park, J., Karplus, K., Barrett, C., Hughey, R., Haussler, D., Hubbard, T., and Chothia, C. (1998) Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *Journal of Molecular Biology* **284**: 1201-1210.

Ponting, C. P., Aravind, L., Schultz, J., Bork, P., and Koonin, E. V. (1999) Eukaryotic signaling domain homologues in Archaea and Bacteria. Ancient ancestry and horizontal gene transfer. *Journal of Molecular Biology* **289**: 729-745.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucleic Acids Research* **35**: 61-65.

Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**: 276-277.

Ritalahti, K.M., and Löffler, F.E. (2004) Populations implicated in anaerobic reductive dechlorination of 1,2-dichloropropane in highly enriched bacterial communities. *Applied and Environmental Microbiology* **70**: 4088-4095.

Sanford, R.A., Cole, J.R., and Tiedje, J.M. (2002) Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68**: 893-900.

Schmidt, H.A., and Haeseler, A.v. (eds) (2003) *Maximum likelihood analysis using TREE-PUZZLE.* New York: Wiley and Sons.

Scholz-Muramatsu, H., Neumann, A., Messmer, M., Moore, E., Diekert, G. (1995) Isolation and characterization of *Dehalospirillum multivorans* gen. nov., a tetrachloroethene-utilizing, strictly anaerobic bacterium. *Archives of Microbiology* **163**: 48-56

Schumacher, W., Holliger, C., Zehnder, A.J.B., and Hagen, W.R. (1997) Redox chemistry of cobalamin and iron-sulfur cofactors in the tetrachloroethene reductase of *Dehalobacter restrictus. FEBS Letters* **409**: 421-425.

Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C., Richter, A.R., and White, O. (2007) TIGRFAMs and genome properties: tools for the assignment of molecular function and biological processes in prokaryotic genomes. *Nucleic Acids Research* **35**: 260-264.

Seshadri, R., Adrian, L., Fouts, D.E., Eisen, J.A., Phillippy, A.M., Methe, B.A. et al. (2005) Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105-108.

Siebert, A., Neumann, A., Schubert, T., and Diekert, G. (2002) A non-dechlorinating strain of *Dehalospirillum multivorans*: evidence for a key role of the corrinoid cofactor in the synthesis of an active tetrachloroethene dehalogenase. *Archives of Microbiology* **178**: 443-449.

Smidt, H., and Vos, W.M.d. (2004) Anaerobic microbial dehalogenation. *Annual Review of Microbiology* **58**: 43-73.

Staub, O. and Rotin, D. (1996) WW Domains. *Structure* **4**: 495-499.
Sung, Y., Ritalahti, K.M., Sanford, R.A., Urbance, J.W., Flynn, S.J., Tiedje, J.M., and

Löffler, F.E. (2003) Characterization of two tetrachloroethene-reducing, acetate-oxidizing anaerobic bacteria and their description as *Desulfuromonas michiganensis* sp. nov. *Applied and Environmental Microbiology* **69**: 2964-2974.

Sung, Y., Ritalahti, K. M., Apkarian, R. P., and Löffler, F.E. (2006) Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring *Dehalococcoides* isolate. *Applied and Environmental Microbiology* **72**:1980-1987.

Sung, Y., Fletcher, K.E., Ritalahti, K.M., Apkarian, R.P., Ramos-Hernández, N., Sanford, R.A. et al. (2006) *Geobacter lovleyi* sp. nov. strain SZ, a novel metal-reducing and tetrachloroethene-dechlorinating bacterium. *Applied and Environmental Microbiology* **69**: 2775-2782.

Suyama, A., Yamashita, M., Yoshino, S., and Furukawa, K. (2002) Molecular characterization of the PceA reductive dehalogenase of *Desulfitobacterium* sp. Strain Y51. *Journal of Bacteriology* **184**: 3419-3425.

Swofford, D.L. (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods) 4.

Thibodeau, J., Gauthier, A., Duguay, M., Villemur, R., Lépine, F., Juteau, P., and Beaudet, R. (2004) Purification, cloning, and sequencing of a 3,5-dichlorophenol reductive dehalogenase from *Desulfitobacterium frappieri* PCP-1. *Applied and Environmental Microbiology* **70**: 4532-4537.

Than, C., Ruths, D., Innan, H., and Nakhleh, L. (2007) Confounding factors in HGT detection: Statistical error, coalescent effects, and multiple solutions. *Journal of Computational Biology* **14**: 517-535.

Thomas, S.H., Wagner, R.D., Arakaki, A.K., Skolnick, J., Kirby, J.R., Shimkets, L.J., Sanford, R.A., Löffler, F.E. (2008) The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* **3:** e2103.

Tian, W., and Skolnick, J. (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of Molecular Biology* **333**: 863-882.

Todd, A.E., Orengo, C.A., and Thornton, J.M. (2001) Evolution of function in protein superfamilies, from a structural perspective. *Journal of Molecular Biology* **307**: 1113-1143.

Tsukagoshi, N., Ezaki, S., Uenaka, T., Suzuki, N., and Kurane, R. (2005) Isolation and transcriptional analysis of novel tetrachloroethene reductive dehalogenase gene from *Desulfitobacterium* sp. strain KBC1. *Applied Microbiology and Biotechnology.* **69**: 543-553.

Utkin, I., Woese, C., and Wiegel, J. (1994) Isolation and characterization of *Desulfitobacterium dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *International Journal of Systematic Bacteriology* **44**: 612-619.

van den Pas, B.A.v.d., Smidt, H., Hagen, W.R., Oost, J.v.d., Schraa, G., Stams, A.J.M., and Vos, W.M.d. (1999) Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in *Desulfitobacterium dehalogenans*. *The Journal of Biological Chemistry* **274**.

Villemur, R., Lanthier, M., Beaudet, R., and Lepine, F. (2006) The *Desulfitobacterium* genus. *FEMS Microbiology Reviews* **30**: 706-733.

Wagner, A., Adrian, L., Kleinsteuber, S., Andreesen, J. R., Lechner, U. (2009) Transcription analysis of genes encoding homologues of reductive dehalogenases in "*Dehalococcoides*" sp. Strain CBDB1 by using terminal restriction fragment length polymorphism and quantitative PCR. *Applied and Environmental Microbiology* **75**:1876-1844.

Wagner, D.D., Hug, L.A., Hatt, J.K., Spitzmiller, M.R., Padilla-Crespo, E., Ritalahti, K.M., Edwards, E.A., Konstantinidis, K.T., and Löffler, F.E. (2012) Genomic determinants of organohalide-respiration in *Geobacter lovleyi*, an unusual member of the *Geobacteraceae. BMC Genomics* **13**: 200.

Yan, J., Rash, B. A., Rainey, F. A., Moe, W. M. (2009) Isolation of novel bacteria within the *Chloroflexi* capable of reductive dechlorination of 1,2,3-trichloropropane. Environmental Microbiology **11**: 833-843.

CHAPTER 4

GENOMIC DETERMINANTS OF ORGANOHALIDE-RESPIRATION IN *GEOBACTER LOVLEYI*, AN UNUSUAL MEMBER OF THE *GEOBACTERACEAE.*

Used with permission:

Wagner[1], D.D., Hug[2], L.A., Hatt[3], J.K., Spitzmiller[3], M.R., Padilla-Crespo[4], E., Ritalahti[4], K.M., Edwards[5], E.A., Konstantinidis[1,3], K.T., and Löffler[4,6,7], F.E. (2012) Genomic determinants of organohalide-respiration in *Geobacter lovleyi*, an unusual member of the *Geobacteraceae. BMC Genomics* 13: 200. Copyright 2012, BioMed Central

[1] School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, GA, 30332, USA

[2] Department of Cell and Systems Biology, University of Toronto, Toronto, ON M5S 3E5, Canada

[3] School of Civil and Environmental Engineering, 311 Ferst Drive, Georgia Institute of Technology, Atlanta, GA, 30332, USA

[4] Department of Microbiology, University of Tennessee, M409 Walters Life Science Bldg, Knoxville, Tennessee 37996, USA

[5] Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, ON M5S 3E5, Canada

[6] Biosciences Division, Oak Ridge National Laboratory, Bethel Valley Road, Building 1520, Oak Ridge, Tennessee 37831, USA

[7] Department of Civil and Environmental Engineering, University of Tennessee, 233 Perkins Hall, Knoxville, Tennessee 37996, USA

4.1 Abstract.

4.1.1 Background.

Geobacter lovleyi is a unique member of the *Geobacteraceae* because strains of this species share the ability to couple tetrachloroethene (PCE) reductive dechlorination to *cis*-1,2-dichloroethene (*cis*-DCE) with energy gain and growth (i.e., organohalide respiration). Strain SZ also reduces U(VI) to U(IV) and contributes to uranium immobilization, making *G. lovleyi* relevant for bioremediation at sites impacted with chlorinated ethenes and radionuclides. *G. lovleyi* is the only fully sequenced representative of this distinct *Geobacter* clade, and comparative genome analyses identified genetic elements associated with organohalide respiration and elucidated genome features that distinguish strain SZ from other members of the *Geobacteraceae.*

4.1.2 Results.

Sequencing the *G. lovleyi* strain SZ genome revealed a 3.9 Mbp chromosome (G + C=54.7%) with average amino acid identities of 53-56% compared to other sequenced *Geobacter* spp. Sequencing also revealed the presence of a 77 kbp plasmid, pSZ77 (G + C=53.0%), with nearly half of its encoded genes corresponding to chromosomal homologs in other *Geobacteraceae* genomes. Among these chromosome-derived features, pSZ77 encodes 15 out of the 24 genes required for *de novo* cobalamin biosynthesis, a required cofactor for organohalide respiration. A plasmid with 99% sequence identity to pSZ77 was subsequently detected in the PCE-dechlorinating *G. lovleyi* strain KB-1 present in the PCE-to-ethene-dechlorinating consortium KB-1. Additional PCE-to-*cis*-DCE-dechlorinating *G. lovleyi* strains obtained from the PCE-contaminated Fort Lewis, WA, site did not carry a plasmid indicating that pSZ77 is not a requirement (marker) for PCE respiration within this species. Chromosomal

genomic islands found within the *G. lovleyi* strain SZ genome encode two reductive dehalogenase (RDase) homologs and a putative conjugative pilus system. Despite the loss of many *c*-type cytochrome and oxidative-stress-responsive genes, strain SZ retained the majority of *Geobacter* core metabolic capabilities, including U(VI) respiration.

4.1.3 Conclusions.

Gene acquisitions have expanded strain SZ's respiratory capabilities to include PCE and TCE as electron acceptors. Respiratory processes core to the *Geobacter* genus, such as metal reduction, were retained despite a substantially reduced number of *c*-type cytochrome genes. pSZ77 is stably maintained within its host strains SZ and KB-1, likely because the replicon carries essential genes including genes involved in cobalamin biosynthesis and possibly corrinoid transport. Lateral acquisition of the plasmid replicon and the RDase genomic island represent unique genome features of the PCE-respiring *G. lovleyi* strains SZ and KB-1, and at least the latter signifies adaptation to PCE contamination.

4.2 Introduction.

Geobacter spp. are common members of anoxic freshwater sediment and subsurface microbial communities, where they are involved in the reduction of oxidized metal species and the turnover of organic matter (Coates et al., 1996). Members of this genus show promise for bioremediation of anoxic subsurface environments contaminated with toxic radionuclides (Anderson et al., 2003). While dissimilatory metal reduction is a hallmark feature of *Geobacter*, the ability to use chlorinated organic compounds as electron acceptors has only recently been discovered in this genus, and appears to be restricted to a single *Geobacter* clade (Fig. S1) with only a few cultured

representatives (Sung et al., 2006). Organohalide respiration has been described for *Geobacter thiogenes* strain K and *Geobacter lovleyi* strain SZ, which dechlorinate trichloroacetate to dichloroacetate and tetrachloroethene (PCE) to *cis*-1,2-dichloroethene (*cis*-DCE), respectively (Sung et al., 2006; DeWever et al., 2000). In addition, another PCE-to-*cis*-DCE-dechlorinating *G. lovleyi* strain, designated strain KB-1, was identified in the PCE-to-ethene-dechlorinating consortium KB-1 (Duhamel and Edwards, 2006). *G. lovleyi* 16S rRNA gene sequences have been detected at the contaminated Oak Ridge IFRC site (Amos et al., 2007) and in trichloroethene (TCE)-contaminated sediments from Ft. Lewis, WA (Costanza et al., 2009). A recent continuous flow column study using the PCE-to-ethene-dechlorinating bioaugmentation consortium Bio-Dechlor INOCULUM (BDI) containing *G. lovleyi* strain SZ indicated that PCE-dechlorinating *Geobacter* strains enhance dissolution of free phase PCE (Amos et al., 2009). Further, *G. lovleyi* strain SZ uses graphite electrodes as a direct electron donor for reductive dechlorination (organohalide respiration), possibly enabling innovative bioremediation approaches (Strycharz et al., 2008). The organohalide-respiring *Geobacter* strains share 16S rRNA gene sequences with 98% to 100% identity to each other but only 93% identity with *G. sulfurreducens* strain PCA, the type strain for the *Geobacter* genus. Among the *δ-Proteobacteria*, *Desulfuromonas michiganensis* is the only other PCE-to-*cis*-DCE-respiring species but genome information is not available (Sung et al., 2003). The genus *Anaeromyxobacter* (*δ-Proteobacteria*) comprises isolates with sequenced genomes, which are capable of using chlorinated phenols as respiratory electron acceptors, but PCE dechlorination has not been reported (Sanford et al., 2002; Thomas et al., 2010; Thomas et al., 2008).

Due to its implications for bioremediation of sites contaminated with both chlorinated ethenes and radionuclides (Sung et al., 2006), and the paucity of genome information of PCE-respiring *δ-Proteobacteria*, *G. lovleyi* strain SZ is a promising reference strain for understanding the physiological and evolutionary responses of microbes to anthropogenic changes in the environment. Here we present the *G. lovleyi* strain SZ genome sequence, including the discovery of a unique 77 kbp plasmid (pSZ77). The plasmid is notable, as it contains laterally acquired genes and some that encode functions important to *Geobacter* metabolism such as cobalamin biosynthesis. Comparative analyses of the strain SZ genome revealed significant functional divergence from previously sequenced *Geobacter* spp. and mobile elements sharing genomic features of *Pelobacter*, a distinct genus of the *δ-Proteobacteria.*

4.3 Results and discussion.

4.3.1 The *Geobacter lovleyi* strain SZ genome.

Sequencing of the *G. lovleyi* strain SZ genome revealed both typical *Geobacteraceae* characteristics (e.g., genes encoding multiheme *c*-type cytochromes) and elements not previously found among members of the *Geobacter* genus, including genes encoding putative reductive dehalogenases (RDases) and a 77-kbp plasmid designated pSZ77. The 54.8% G + C content, 3,644 predicted open reading frames (ORFs), and 3.9 Mb size of the strain SZ chromosome (Table 1) are comparable to other sequenced genomes of members of the *Geobacter* genus, *G. sulfurreducens* PCA (Methe et al., 2003), *G. metallireducens* GS-15 (Aklujkar et al., 2009), *G. bemidjiensis* Bem (Aklujkar et al., 2010), *G. uraniireducens* Rf4 (Shelobolina et al., 2008), and *Geobacter* spp. strains FRC-32 (Prakash et al., 2010), M21, and M18, all ranging from

3.8-5.1 Mb in size. The strain SZ chromosome has 279 chromosomal genes assigned to the energy production clusters of orthologous groups (COGs) (class C), a somewhat lower count compared to total COGs class C genes on the seven other sequenced *Geobacter* genomes (ranging from 280 to 356, avg. 334), indicating a shift in the strain SZ respiration-related gene repertoire. By comparison, the 481 strain SZ genes functionally classified in the signal transduction COGs (class T) lies within the range for *Geobacter* spp. (382 to 587, avg. 427). SZ chromosomal genes have 40% of their top BlastP matches in the genomes of other *Geobacter* spp. (Table 1), where average BlastP identities between the strain SZ and each *Geobacter* proteome range from 53-56%. By comparison, pSZ77 has 53.0% G + C content and only eight (11%) of its 81 total predicted ORFs with top BlastP matches (avg. identity 77%) among *Geobacter* spp. (Table 2).

4.3.2 Organohalide respiration.

A defining feature distinguishing the *G. lovleyi* strain SZ chromosome from other *Geobacter* genomes is the presence of a gene cluster related to organohalide respiration. The ability of *G. lovleyi* strains to respire PCE resides on a chromosomal genomic island (Figure 1A) containing two putative PceA RDases. The *pce*-genes predicted to play a role in PCE respiration, *pceT-pceC-pceA1-pceB1-pceA2-pceB2*, (Glov_2866-Glov_2875) comprise a region with a G + C content of 37%, as much as 2.5 standard deviations below the chromosomal average of 54.8%. The Codon Adaptation Index (CAI), a comparative measure of codon usage, for the six *pce*-genes cluster fall below the average for SZ chromosomal ORFs (normalized CAI > 1.00, Table S1), indicating recent acquisition by the SZ genome. The 'Pce' chromosomal region, in which the *pce*-genes reside, exhibits an apparent reversal in GC-skew

(Figure S2). Finally, the six *pce*-genes have no homologs in any other *Geobacter* or related *Pelobacter* genomes. Together, these features indicate that the SZ chromosomal Pce region encoding the *pce*-genes in an atypical region acquired by lateral gene transfer.

The six *pce*-genes of the *G. lovleyi* genome are homologous to genes in functionally characterized PCE respiration gene clusters. Functions for three of the four components of the *pce*-gene cluster, *pceA-pceB-pceC-pceT*, from the *Firmicutes* (Figure 1B) is supported by experimental evidence in *Desulfitobacterium hafniense* strains TCE1 (Maillard et al., 2005) and Y51 (Furukawa et al., 2005) and *Dehalobacter restrictus* strain PER-K23 (Maillard et al., 2003). *pceA* encodes the catalytic PceA RDase subunit, whose activity towards organohalides is dependent upon a bound cobalamin cofactor (Maillard et al., 2003). *pceB* is inferred to encode a membrane-bound subunit to PceA, but direct experimental evidence for this function is lacking. *pceC* is co-transcribed with *pceA-pceB* in *D. hafniense* Y51 and the PceC protein is believed to function in regulating *pceA* gene expression and electron transfer to the PceA protein (Futagami et al., 2006). *pceT* encodes a protein shown to function as a chaperone to the PceA preprotein (Morita et al., 2009). The order of the predicted *G. lovleyi pce*-genes, *pceT-pceC-pceA1-pceB1-pceA2-pceB2*, differs from that of the functionally-characterized *pce*-gene clusters (Figure 1B). The *G. lovleyi pceA-pceB* ORFs, encoding the predicted PCE RDase catalytic subunit and the membrane anchor subunit, respectively, are duplicated on tandem 2,466 bp long blocks sharing 99.8% nucleotide identity (Figure 1A). Despite apparent differences in organization, the amino acid sequences encoded by all six *G. lovleyi pce*-genes have their most similar homologs among the functionally characterized *pce*-gene clusters from *Desulfitobacterium hafniense* strain Y51 (AB070709), *D. hafninese* strain TCE1

(AJ439608), and *D. restrictus* (AJ439607). Both sets of strain SZ's putative PceA and PceB proteins share 33-36% amino acid (aa) identity with PceA (RDase A subunit, BAE84628) and PceB (RDase B subunit, BAE84627) from *D. hafniense* Y51 (Furukawa et al., 2005). The *G. lovleyi* PceAs share only distant similarity with the genomic-island-encoded VcrA RDase of *Dehalococcoides* sp. strain VS (18% identity, 30% similarity) (McMurdie et al., 2007). Glov_2869 (putative PceC) shares a FMN-binding domain (pfam04205), a polyferredoxin (COG0348) domain, and 33% aa identity with the *D. hafniense* Y51 PceC regulatory protein (BAE84626). Glov_2868 (putative PceT) is annotated as 'peptidylprolyl isomerase' and shares 19% aa identity with PceT from *D. hafniense* (BAE84625). Despite their apparent shared conserved functions, the low sequence similarities and lack of synteny between strain SZ and the *Desulfitobacterium* spp. *pce*-gene clusters suggests the gene clusters have diverged from their shared ancestral gene cluster over time.

An IS21-like integrase gene cluster and a transposase gene flank the strain SZ *pce*-genes upstream and downstream, respectively, but neither exhibits the compositional features suggestive of recent lateral gene transfer, with G + C content, codon biases, and GC-skew consistent with the *G. lovleyi* genome. Transposase-associated repeats were found to mediate circularization, and presumably excision, of the *D. hafniense* strain TCE1 PceA-encoding transposon (Maillard et al., 2005). In contrast, no repeats flanking the SZ PceA genomic island could be detected, and the ISL3-superfamily transposase (pfam01610) downstream of the SZ *pce*-genes lacks detectable similarity with the mutator family transposases (pfam00872) found on the *D. hafniense* strain TCE1 transposable element. Furthermore, the *rve* family integrase (pfam00665) and IS21-type ATPase/integrase cluster upstream of the SZ *pce*-genes, along with adjacent non-coding DNA (Figure 1A), do not share homology with the

dsiB integrase in the *vcrA* genomic islands of *Dehalococcoides* (McMurdie et al., 2011) nor any other mobilization-related genomic elements (inverted repeats, direct repeats, etc.) associated with RDase genes (McMurdie et al., 2009). The integrase and the ATPase instead appear to originate from within the *Geobacter* genus, sharing 62% and 80% aa identity, respectively, with their homologs in *G. uraniireducens*. The SZ Pce genomic island integrase and transposase lack homology with known RDase-associated mobile elements, while the G + C content of the *pce*-genes suggests much of this element was acquired by strain SZ relatively recently. The apparent divergence of the strain SZ *pce*-genes from their closest known orthologs in *Desulfitobacterium* spp. suggests the SZ genes are not a recent lateral transfer from the *Desulfitobacterium*/*Dehalobacter* group but from another, yet unidentified donor.

4.3.3 F-factor conjugation.

A second predicted chromosomal genomic island harbors a predicted F-factor conjugative pilus *tra*-gene cluster (Glov_0304 - Glov_0322), *traE-L-E-K-B-V-C-N-N-W, traU-trbC, traF, traG,* (Figure 2), which lacks homologs in any sequenced *Geobacter* genome. The cluster of *tra*-genes form part of a region ('Tr' in Figure S2) spanning 108 ORFs, flanked by a transposase (Glov_0226) and a resolvase (Glov_0336) gene. Aside from the 19 *tra*-genes, 20 of the Tr region genes are assigned to COGs class L or class V, which include endo/exoribonucleases and helicases, while another 42 genes have hypothetical or unknown functions. A total of 62 genes in the Tr region, including a majority of the *tra*-genes, share > 75% aa identity with orthologs on the chromosome of *Pelobacter propionicus* DSM 2380. Given that the average SZ chromosome-encoded protein shares 57% identity with their BlastP matches in *P. propionicus*, the Tr region composition suggests lateral acquisition of

this region, possibly from a *Pelobacteraceae*-like donor. The region is designated as a genomic island based on its similarity to a non-*Geobacteraceae* genome and the predominance of hypothetical genes and DNA-manipulating genes (Merkl, 2006). An alternate hypothesis for the origin of the Tr region is a shared ancestry within the order *Desulfuromonadales*, which includes the *Pelobacteraceae* and the *Geobacteraceae*, with subsequent loss of this element in the *Geobacter*. With only seven sequenced *Geobacter* genomes available, vertical inheritance cannot be ruled out; however, the high similarity to homologs in *Pelobacter* does suggest that the *tra*-gene cluster is part of a genomic island.

The *tra*-gene cluster, Glov_0304 to Glov_0322, encodes proteins homologous to known plasmid-encoded DNA-transfer F-type conjugative pili (Lawley et al., 2003b). Out of the 19 proteins encoded on the *tra*-gene cluster, five have PSI-Blast matches to genes unique to F-type conjugative pili (i.e., *traN*, *traW*, and *traU*), while seven have PSI-Blast matches to conserved F-type and P-type conjugative pilus "core" genes: *traE*, *traL*, *traK*, *traB*, *traV*, *traC*, and *traG* (Lawley et al., 2003b; Lawley et al., 2003a) (Figure 2). All but one of the proteins on the strain SZ putative F-type pilus cluster share synteny and high similarity with genes on the chromosome of *P. propionicus* (sharing 40% to 91% aa identity) and the 202 kbp plasmid pPRO1 of *P. propionicus* (37% to 61% aa identity) (Figure 2). Nearly half of the 19 genes in the strain SZ conjugative pilus cluster have G + C contents near the strain SZ genomic average and normalized CAI > 1.00 (Table S2), suggesting a sufficient residence time on the SZ genome to ameliorate to SZ chromosomal codon usage (Lawrence and Ochman, 1998). Four additional genomic islands on the SZ chromosome are predicted based on low G + C%, disruption of GC-skew, and/or the lack of homologs in a majority of other *Geobacter* spp. genomes. These additional inferred genomic islands,

hyp1, hyp2, and hyp3 encode hypothetical proteins and proteins assigned to COGs class L or class V, while genomic island M contains five genes predicted to function in capsule polysaccharide biosynthesis (Figure S2).

4.3.4 Electron transfer and oxidoreductases.

Additional features distinguishing the *G. lovleyi* strain SZ genome from other *Geobacteraceae* include the reduced number of genes encoding *c*-type cytochromes and the lack of several key genes related to oxygen tolerance and reactive oxygen species (ROS) detoxification. The *G. lovleyi* strain SZ genome encodes 49 ORFs encoding *c*-type cytochromes (Table 3, Table S3), of which only six are predicted multi-heme cytochromes with 10 or more CxxCH heme-binding motifs, and none have more than 12 CxxCH motifs (Table 3). In contrast, described *Geobacter* genomes encode 76–104 *c*-type cytochromes, of which 17–31 have 10 or more CxxCH motifs, with a maximum of 43 heme-binding motifs (Butler et al., 2010). *P. propionicus* has 20 ORFs predicted to encode *c*-type cytochromes, with only one having more than 10 heme-binding motifs (Table 3, Table S4).

The reduced set of multi-heme *c*-type cytochromes in strain SZ (Table 3, Table S3) appears adequate in mediating respiration with U(VI), Mn(IV), and Fe(III) oxides (Sung et al., 2006) and electrode surfaces (Strycharz et al., 2008). Specific functions have only been determined for cytochromes with 10 or fewer hemes (Butler et al., 2010). Decaheme *c*-type cytochromes have been confirmed to mediate respiration on electrodes in *Shewanella* (El-Naggar et al., 2010) and an 8-heme outer-membrane cytochrome has been shown to be involved in electron transfer to solid state electrodes in *Geobacter sulfurreducens* (Inoue et al., 2011). Smaller outer membrane *c*-type cytochromes with 4 to 6 hemes are essential for *Geobacter* respiration with insoluble

Fe(III) and Mn(IV) oxides (Mehta et al., 2005). PilA-type pilins may play an essential role in metal reduction (Reguera et al., 2005), particularly in radionuclide reduction (Cologgi et al., 2011). Strain SZ encodes a putative PilA (Glov_2096) sharing 81% aa identity with the conductive pilus protein (PilA) of *G. sulfurreducens* (GSU1496) (Reguera et al., 2005). *P. propionicus* also encodes a putative PilA (Ppro_1656), which shares 80% identity with the *G. sulfurreducens* PilA. Yet, *P. propionicus* has a reduced repertoire of *c*-type cytochromes, and is completely lacking decaheme *c*-type cytochromes (Table S4). Accordingly, *P. propionicus* lacks the ability to respire on electrodes or radionuclides (Butler et al., 2009). By contrast, the strain SZ genome carries the predicted minimal set of genes to allow utilization of a comparable range of electron acceptors as other *Geobacter* spp. (Sung et al., 2006).

Additional genes that are key to *Geobacter* electron transfer are present as multiple paralogs on the strain SZ genome. The strain SZ genome encodes 11 molybdopterin oxidoreductase domain proteins (pfam00384), a larger number compared to other *Geobacter* genomes. The strain SZ molybdopterin oxidoreductase-type proteins, for which function can be inferred, include three nitrate reductases and two formate dehydrogenases (Table S5). One of the nitrate reductase proteins is a periplasmic-type, sharing 43% aa identity (61% similarity) with the respiratory nitrate reductase (NapA) from *Desulfovibrio desulfuricans* (Marietou et al., 2005). The role of the two predicted formate dehydrogenases is unclear, as strain SZ does not utilize formate as an electron source or carbon source under PCE- or Fe(III)-reducing conditions (Sung et al., 2006). The SZ chromosome encodes seven fumarate reductase/succinate dehydrogenase-type flavoproteins (pfam00890), while no other *Geobacter* genome encodes more than three. Strain SZ uses fumarate as an electron acceptor and has the full set of genes encoding TCA cycle enzymes, but it is not clear

which of these fumarate reductases play a role in these pathways. Strain SZ has five gene clusters (one on pSZ77) encoding pyruvate ferredoxin/flavodoxin oxidoreductase (PFOR) complexes (Table S5) of the type inferred to mediate incorporation of acetate into the TCA cycle in *G. sulfurreducens* (Mahadevan et al., 2006). Strain SZ lacks any homologs to subunit A of the pyruvate dehydrogenase E1 complex, making it unclear how the organism is able to use pyruvate as an electron donor (Sung et al., 2006).

With regard to ROS and oxygen detoxification, strain SZ lacks homologs to the heme-copper (*aa3*) cytochrome *c* oxidases involved in the utilization of oxygen in *G. sulfurreducens* (Nuñez et al., 2006). Accordingly, strain SZ does not respire oxygen (Sung et al., 2006) and responds negatively to in situ oxygen exposure (Thomas et al., 2010). Strain SZ possesses a complete gene cluster encoding homologs to the cytochrome *bd* ubiquinol oxidase complex shown to mediate oxygen tolerance in anaerobes (Glov_1208-Glov_1209, Table S6) (Das et al., 2005), but lacks any homolog to the diheme cytochrome *c* peroxidase (MauG) or superoxide dismutase (SodA) shown to be strongly expressed in the presence of oxygen in *G. uraniireducens* (Mouser et al., 2009). *P. propionicus* similarly lacks heme-copper (*aa3*) cytochrome *c* oxidase, MauG, and SodA. The absence of multi-heme *c*-type cytochromes with more than 12 hemes in strain SZ may reflect differences in habitat and metabolic strategy as compared to other *Geobacter* spp. The high-molecular weight *c*-type cytochromes with 13 or more hemes in *G. sulfurreducens* were suggested to function as biological capacitors (Lovley, 2008), providing storage for reducing equivalents while the energy-stressed *Geobacter* cell is seeking new sources of terminal electron acceptors. Although this is an intriguing hypothesis, the synthesis of such high molecular weight *c*-type cytochromes under nutrient limitations burdens the cell's energy budget, and the true benefit for survival has yet to be demonstrated. The poor oxygen tolerance

exhibited by strain SZ, confining it to persistently anoxic habitats, may be associated with the reduced count of high molecular weight, multi-heme *c*-type cytochromes prevalent in aerotolerant *Geobacter*. Strain SZ may utilize a strategy for temporary electron storage that is not dependent on *c*-type cytochromes or is able to employ cytochromes with fewer hemes to the same effect seen in aerotolerant *Geobacter*.

Consistent with all other *Geobacteraceae* genomes, strain SZ possesses the full set of genes for the *de novo* biosynthesis of both heme and cobalamin (Table 4). Yet, unlike any other *Geobacter* spp. or *Pelobacter* spp. genome, a majority of cobalamin biosynthesis genes in the SZ genome are localized to a plasmid.

4.3.5 Plasmid pSZ77.

The *G. lovleyi* genome sequence revealed the presence of a 77,113 bp circular plasmid, designated pSZ77. Out of the 81 total predicted plasmid-encoded ORFs, 32 were inferred to be involved in plasmid maintenance and stability along with cobalamin biosynthesis, membrane transport, or other functions likely to contribute critical functions to strain SZ's metabolism (Figure 3). To date, *G. metallireducens* GS-15 and *P. propionicus* are the only closely related genomes to *G. lovleyi* that contain plasmids. The strain GS-15 plasmid is 14 kbp, and encodes genes predicted to be involved strictly in plasmid maintenance (Aklujkar et al., 2009). *P. propionicus* has a multipart genome comprised of a 4.0 Mbp chromosome and two plasmids. The larger 202 kbp plasmid pPRO1 has COGs class functions related to energy production, and encodes the two cytochrome *c* ammonia-forming nitrite reductases, NrfA and NrfH, and a *c*-type cytochrome of unspecified function. The smaller 31 kbp plasmid pPRO2 encodes two additional *c*-type cytochromes (Table S4). No *c*-type cytochromes are

encoded on pSZ77, but 18% of the pSZ77 genes have predicted COGs class functions in coenzyme metabolism, mostly cobalamin biosynthesis (Table 2, Figure S3).

4.3.6 Plasmid maintenance.

Several pSZ77 protein-coding genes have putative function in plasmid replication or segregational stability and are associated with a predicted origin of replication, *oriR* (Figure 4). Glov_3681 encodes a protein with the replication initiation protein (RepA) domain (pfam04796). pSZ77 RepA may function in a complex with the SZ-chromosome-encoded DnaA protein to initiate plasmid replication (Sharma et al., 2003; del Solar et al., 1998) at predicted DnaA-binding sites (Schaper and Messer, 1995) within the predicted *oriR* downstream of *repA* on pSZ77 (Figure 4). The pSZ77 RepA appears to share phylogenetic affiliation with the RepA of various plasmids from *β*- or *γ*-*Proteobacteria* (Figure 5), including an IncQ-like mobilizable plasmid (Doublet et al., 2009) and an IncP-1-like environmental plasmid (Mela et al., 2008). Yet, pSZ77 RepA does not share more than 39% aa identity (57% similarity) with the RepA encoded by either of these two characterized plasmids nor any other homologs in the public databases, including its two homologs from the plasmids of *δ*-*Proteobacteria* (identity 33% to 35%; similarity 49% to 55%). Upstream of pSZ77 *repA*, Glov_3684 encodes a ParA-type plasmid partitioning ATPase (Table S7) and may provide a mechanism for distribution of pSZ77 copies to daughter cells after division (Hayes, 2000). Glov_3687-Glov_3688 encode proteins homologous to HipB and HipA (Figure 4), a putative toxin-antidote system (Table S7), which may further contribute to pSZ77 stability (Nordström and Austin, 1989). Together, these observations suggest that the pSZ77 replicon belongs to a new plasmid incompatibility

group with inferred mechanisms for both active partitioning and post-segregational stability.

4.3.7 Plasmid gene clusters associated with cobalamin metabolism.

The strain SZ chromosome contains several genes encoding cobalamin-dependent enzymes including ribonucleotide reductase, methionine synthase, and methylmalonyl-CoA mutase; genes that are found in all other *Geobacter* genomes (Table S8). In addition, the strain SZ chromosome encodes two PCE RDases that presumably require a cobalamin co-factor (Maillard et al., 2003). Accordingly, strain SZ has genes encoding all inferred functions necessary for *de novo* cobalamin biosynthesis plus a predicted outer membrane transport system, which may play a role in uptake of extracellular corrinoids. Unlike other *Geobacter* genomes, the majority of strain SZ cobalamin biosynthesis genes are encoded on the plasmid rather than the chromosome, localized in two gene clusters and a single gene locus (Table 4). The first cobalamin biosynthesis gene cluster comprises 15 consecutive or overlapping genes (Glov_3646 through Glov_3660) together encoding enzymes mediating the first 11 steps of *de novo* cobalamin biosynthesis and a cobalt ABC transporter. Genes 2 through 10 in the first cluster, *cbiJ-H-G-F-D-L-K-C-A-CysG*, along with one of the cobalt transport genes, *cbiN*, lack isofunctional homologs on the SZ chromosome suggesting the plasmid is essential when extracellular sources of cobalamin are lacking. The second cobalamin biosynthesis gene cluster comprises *cobT* (Glov_3678) on the reverse strand and *cobD*/*cbiP* (Glov_3679) on the leading strand, both showing evidence of duplication with genes on the SZ chromosome. CobT shares 78% aa identity with a chromosome-encoded CobT homolog. Similarly, the CobD/CbiP fusion shares 77% identity with a homologous fusion protein encoded on the SZ chromosome. The third

locus is comprised of a single gene, *cobA* (Glov_3718), possibly involved in the adenosylation step of cobalamin biosynthesis and CobA shared 36% aa identity with CobA encoded on the SZ chromosomal *cobA* locus. Genes encoding four of the final five steps of cobalamin biosynthesis, *cbiB*, and *cobU*, *cobS*, and *cobC*, are found exclusively on the SZ chromosome (Table 4).

Nine genes highly similar to chromosomal orthologs are inferred in outer membrane transport and are situated between both cobalamin biosynthesis gene clusters on pSZ77 (Figure 3). Proteins encoded by genes at the 5' and 3' ends of this cluster, Glov_3667 and Glov_3675, were inferred by Pred-TMBB (Bagos et al., 2004) to have up to 20 trans-membrane ß-strands characteristic of the TonB-dependent outer membrane receptors (Ferguson and Deisenhofer, 2002). Proteins encoded by Glov_3670, Glov_3669, and Glov_3668 have sequence characteristics matching the ExbB, ExbD, and TonB proteins involved in outer membrane transport of iron siderophores or vitamin B_{12} (Ferguson and Deisenhofer, 2002; Higgs et al., 1998). Glov_3671 encodes a protein annotated as a cobalt chelatase (CobN) (Figure 3) of the type involved in late cobalt insertion (i.e., "aerobic") cobalamin biosynthesis (Scott and Roessner, 2002). The confirmed CobN cobalt chelatase of *Pseudomonas denitrificans* is associated with pCobS and pCobT subunits (Debussche et al., 1992). No genes predicted to encode homologs to these pCobS and pCobT accessory proteins were identified on the plasmid or chromosome of strain SZ, suggesting the pSZ77 CobN has a distinct function from its homolog in *P. denitrificans*. The pSZ77 putative CobN shared high aa similarity (79% identity) with a protein encoded by Gura_0774 on the *G. uraniireducens* chromosome, which is flanked by *tonB* receptor-related genes. These observations, coupled with its proximity to cobalamin biosynthesis clusters, suggest that the pSZ77-encoded CobN is associated with cobalamin

metabolism in strain SZ, though its exact biochemical role is unknown. The *cobN-exbB-exbD-tonB* gene cluster matches chromosomal homologs in only two related genomes, *G. uraniireducens* and *P. propionicus*, and is predicted to function in the transport of iron siderophores, vitamin B_{12}, or corrinoid precursors (Braun, 1995; Higgs et al., 1998; Ferguson and Deisenhofer, 2002).

pSZ77 genes encoding cobalamin biosynthesis and cobalamin transport functions suggest that the plasmid might be essential for strain SZ growth. To examine the propensity for pSZ77 loss from *G. lovleyi* strain SZ, a series of protocols developed for curing non-essential plasmids from host organisms were assayed (El-Mansi et al., 2000; Ramesh et al., 2001). SDS at concentrations above 0.1% inhibited growth of strain SZ, but growth was stable over 17 consecutive transfers with ≤0.018% SDS. L-ascorbate at concentrations of up to 10 mM did not affect growth over at least nine consecutive transfers. Strain SZ grew with PCE or fumarate as electron acceptor over at least 40 consecutive transfers in cyanocobalamin-free and cyanocobalamin-amended medium. In samples from all treatments that allowed growth, PCR with the *repA*-targeted primer pair yielded amplicons of the expected size suggesting that the plasmid carrying the *repA* DNA fragment was maintained under the growth conditions tested. The observation that strain SZ grew in medium with or without cyanocobalamin amendment indicated that the cobalamin biosynthesis genes on pSZ77 are functional and that pSZ77 is maintained under growth conditions that do not require cobalamin biosynthesis. Our efforts to cure strain SZ of pSZ77 were unsuccessful due either to pSZ77 genes essential to SZ metabolism or due to the effectiveness of plasmid maintenance/stability-related genes. For example, the pSZ77 *cobN*-associated TonB-type outer-membrane transport genes, which match chromosomal orthologs in *G. uraniireducens* and *P. propionicus*, may have been

critical to strain SZ metabolism in the vitamin B_{12}-amended cultures. The pSZ77 *tonB*, *exbB*, *exbD* genes associated with *cobN* do not correspond to a complete cluster on the SZ chromosome, suggesting that transport of cobalamin and/or other organometallic complexes made the plasmid essential under the cultivation conditions used. With these apparently essential functions, pSZ77 shares one of the defining traits of the significantly larger secondary replicons observed in *α-* and *β-Proteobacteria* termed "chromids" (Harrison et al., 2010). Alternatively, the pSZ77 genes predicted to contribute to plasmid stability, such as the partitioning ATPase (*parB*) or the toxin-antitoxin cluster (*hipB-hipA*) may be effective at maintaining the plasmid.

4.3.8 Origins of pSZ77 genes and horizontal gene transfer.

Sequence analyses of pSZ77 genes involved in plasmid maintenance or recombination, cobalamin biosynthesis, and amino acid biosynthesis suggested that a significant portion of the plasmid originated from outside the *Geobacteraceae.* The protein encoded by the putative *hipA* (Glov_3688) shares 75% aa identity with a homolog encoded on the large plasmid (pPRO1) of *P. propionicus.* The pSZ77 replication protein, RepA, shows only distant phylogenetic relatedness with its *Geobacter* and *Pelobacter* homologs (Figure 5) and the *repA* locus has an atypical codon usage (normalized CAI < 1.00, Table S9) for both the SZ chromosome and plasmid. The pSZ77 RepA protein is most similar to homologs of *γ-Proteobacteria* (Figure 5), such as the RepA encoded on plasmid pBI1063 (36% aa identity, 56% aa similarity) from *Stentrophomonas maltophilia.* The predicted plasmid-partitioning gene, *parA*, lacks *Geobacteraceae* or *Pelobacteraceae* homologs altogether (Table S7). Additionally, Glov_3721, encoding a predicted vitamin B_{12}-independent methionine synthase

(MetE), lacks homologs in the *Geobacteraceae* and is functionally redundant to the *metH* locus on the SZ chromosome (Table S8).

The pSZ77 cobalamin biosynthesis genes correspond to isofunctional homologs on the chromosomes of all *Geobacter* spp., but gene fusions and gene order within the cobalamin biosynthesis clusters on pSZ77 are uncharacteristic of cobalamin biosynthesis gene clusters in any other *Geobacter* genome. For instance, the gene fusions *sirA*/*cysG* in the first cobalamin biosynthesis cluster and *cobD*/*cbiP* in the second cluster fully align with their orthologs on the *Pelobacter carbinolicus* chromosome, but share similarity with unfused genes encoded at distinct ORFs in other *Geobacter* genomes. The proteins encoded by genes of the first pSZ77 cobalamin biosynthesis cluster share an average of 56% aa sequence identity with their orthologs on the *P. carbinolicus* chromosome, higher than the average pairwise identity (46%) between the proteomes of strain SZ and *P. carbinolicus*. However, most of the genes in the first and second cobalamin biosynthesis clusters have normalized CAIs > 1.00 (Table S9) suggesting the portion of pSZ77 encoding the cobalamin biosynthesis genes have resided in the SZ genome for sufficient time to ameliorate to its average codon usage (Lawrence and Ochman, 1998; Puigbò et al., 2008).

4.3.9 Distribution of SZ-like plasmids among dechlorinating *Geobacter* strains.

A *G. lovleyi* strain sharing 99% 16S rRNA gene identity with strain SZ was isolated from the PCE-to-ethene-dechlorinating consortium KB-1 in order to examine its plasmid complement. This *Geobacter* isolate was found to be capable of dechlorinating PCE to *cis*-DCE, and was designated *G. lovleyi* strain KB-1. Metagenome sequencing of the KB-1 consortium and plasmid isolation and

sequencing from the *Geobacter* strain KB-1 isolate revealed that strain KB-1 carries a 77,215 bp plasmid sharing high similarity with pSZ77 (99% identity). Although they were originally isolated from different locations and habitats (i.e., strain SZ from non-contaminated Su-Zi Creek freshwater sediment, South Korea (Sung et al., 2006), and strain KB-1 from a chlorinated solvent-contaminated aquifer in Ontario, Canada (Duhamel and Edwards, 2006)), the PCE-to-*cis*-DCE dechlorinating *G. lovleyi* strain KB-1 and strain SZ share highly similar plasmids. The sequence differences between pSZ77 and the KB-1 assembly were comprised mainly of short insertions and deletions. An intact ORF encoding a recombinase on the KB-1 plasmid assembly corresponds to a pseudogene (Glov_3699) with a 119 bp deletion on pSZ77. A 136 bp insertion occurs on pSZ77 of strain SZ in an intergenic region upstream of Glov_3713 encoding a transposase. There are six apparent deletions in the KB-1 assembly relative to the pSZ77 DNA sequence, ranging in length from 1 bp to 136 bp (Table S10), but the majority of these deletions occur in positions in the assembly with two-fold sequence coverage or less, and so cannot be considered reliable. Out of the 79 intact ORFs shared between pSZ77 and the KB-1 assembly, only two single nucleotide polymorphisms could be found. Apparently, the pSZ77 and the KB-1 plasmid assemblies exhibit few sequence differences and little difference in gene repertoire. Likewise, the strain SZ genome shares ~99% identity with metagenome contigs obtained from consortium KB-1. Based on the geographically- and ecologically-disparate habitats of strain SZ and strain KB-1, the strictly anaerobic, non-spore-forming species *G. lovleyi* appears to be cosmopolitan.

The identification of the plasmid from strain KB-1 suggested that the 77 kbp plasmid is a shared feature of members of the organohalide-respiring *Geobacter* clade. To test this hypothesis, primers targeting the *repA* plasmid maintenance gene were

used to amplify the pSZ77 *repA* gene from members of the organohalide-respiring *Geobacter* clade. Included in the analysis were four additional *G. lovleyi* isolates obtained from the chlorinated ethene-contaminated site, Fort Lewis, whose 16S rRNA gene sequences suggest a close phylogenetic relationship (Fig. S1) with strain SZ (99%-100% nucleotide identity). Amplicons from the *repA*-targeted primers of the expected size and sequence were obtained with template DNA from strain SZ (positive control), strain KB-1 (not shown), and the BDI and KB-1 consortia (Fig. S4). No amplicons were obtained from any of the other organohalide-respiring *Geobacter* isolates, suggesting that the Fort Lewis soil *G. lovleyi* isolates Geo7.1, Geo7.2, Geo 7.3, and Geo7.4 do not contain a pSZ77-like plasmid. The primer set Geo564F/840R (Cummings et al., 2003) was successfully used to amplify *Geobacteraceae* 16S rRNA genes from all samples. To further verify that the Fort Lewis *Geobacter* isolates lack plasmids, plasmid extractions were performed with biomass from the Fort Lewis *G. lovleyi* isolates, *G. thiogenes*, and *G. lovleyi* strain SZ. Only strain SZ DNA yielded the characteristic 77 kbp plasmid, and no bands indicative of a plasmid were observed in the other samples (data not shown). The lack of physical DNA in plasmid DNA extractions from *G. thiogenes* and the Fort Lewis *G. lovleyi* strains were consistent with the *repA*-targeted PCR results, suggesting neither *G. thiogenes* nor the Fort Lewis strains contain a similar plasmid. The lack of evidence for pSZ77-like plasmids in several PCE-respiring *Geobacter* suggested that pSZ77 does not play a direct or obligate role in organohalide respiration. Mechanisms for cobalamin biosynthesis or transport of corrinoids are essential for respiration with PCE and other chlorinated ethenes (Maillard et al., 2003). This implies that the *Geobacter* strains without a plasmid likely either have chromosome-encoded cobalamin biosynthesis genes or an efficient mechanism for uptake of extracellular corrinoids. While not indispensable to

reductive dehalogenation across all *G. lovleyi* strains, pSZ77 may play an indirect adaptive role in PCE respiration in strains SZ and KB-1. For instance, pSZ77 may be important to strains SZ and KB-1 by allowing increased gene copy numbers or differential regulation of gene functions relevant for organohalide respiration (Harrison et al., 2010). The two cobalamin biosynthesis clusters on pSZ77 situated on either side of the TonB-encoding transport cluster (Figure 3) may thus provide strains SZ and KB-1 with a mechanism for simultaneous regulation of both cobalamin biosynthesis and scavenging for extracellular corrinoids.

4.4 Conclusions.

Phylogenetic analysis and sequence features related to respiratory metabolism support the affiliation of strain SZ within the *Geobacter* genus. Compared to other *Geobacter* genomes, strain SZ carries an unusually large quantity of laterally acquired DNA including the genomic island with duplicated *pceAB* gene clusters enabling strain SZ to respire PCE and TCE. The origin of the *pce*-genes is unclear given that they share low similarities with their top BlastP matches in *Desulfitobacterium* spp. The frequency and mechanism of transfer of the Pce genomic island are not yet understood, and as such the consequences of potential lateral transfer to non-dechlorinating *Geobacter* spp. are unclear; however, such events are obviously relevant for bioremediation. Up to 69% of strain SZ genes shared their top BlastP matches with other *Geobacter* spp. or *Pelobacter* spp. genomes (Table 1). At the same time, 9% of strain SZ genes did not have homologs in the sequenced representatives of the *δ-Proteobacteria* phylum. Notably, the SZ conjugative pilus *tra*-gene cluster comprises a portion of the largest chromosomal genomic island yet described in a

Geobacter genome. The *tra*-genes may provide strain SZ with a mechanism for enhanced uptake of foreign DNA. Compared to other *Geobacter* genomes, strain SZ possesses fewer *c*-type cytochrome genes with no more than 12 heme binding motifs, which does not appear to limit the organism's ability to respire oxidized metal species.

A distinguishing feature of the *G. lovleyi* genome is pSZ77, an extrachromosomal element carrying genes that are typically located on the chromosome in other *Geobacteraceae.* The phylogenetic origins of the pSZ77 *repA* and *parA* genes are unclear and may have been laterally acquired separately from the pSZ77 genes with chromosomal orthologs. Most notable among its genes with predominantly chromosomal homologs, pSZ77 encodes 15 out of the 24 genes required for *de novo* cobalamin biosynthesis. Although a nearly identical plasmid occurs in *G. lovleyi* strain KB-1, not all organohalide-respiring *Geobacter* spp. carry a non-chromosomal element, indicating that pSZ77 is not a marker for PCE reductive dechlorination within this species. On the other hand, unique genes and gene clusters, particularly the *pce*-genes, may provide discriminative detection of PCE-respiring *Geobacter* strains.

4.5. Methods.

4.5.1 Cultures and growth conditions.

G. lovleyi strain SZ (Sung et al., 2006), *G. lovleyi* strain KB-1 (AY780563) (Duhamel and Edwards, 2006), *G. thiogenes* (DeWever et al., 2000), and four *G. lovleyi* isolates Geo7.1, Geo7.2, Geo7.3, and Geo7.4 obtained from PCE-to-*cis*-DCE-dechlorinating microcosms established with Fort Lewis, WA, soil (Fletcher et al., 2011) were utilized in this study. *G. lovleyi* strain KB-1 was isolated from consortium KB-1. Following a series of eight 1:20 vol/vol dilution transfers to defined mineral medium supplemented

with 90 mg L^{-1} PCE and 10 mM acetate, the final dilution culture was used to inoculate a set of serial dilution agarose shake tubes amended with 10 mM fumarate and 10 mM acetate (Löffler et al., 2005). Transfer of a colony from a 10^{-2} dilution tube to liquid culture yielded a pure culture of *G. lovleyi* strain KB-1. In addition, the PCE-to-ethene-dechlorinating consortia Bio-Dechlor INOCULUM (BDI) and KB-1 (Duhamel and Edwards, 2006) were grown and maintained with 0.05 mg mL^{-1} PCE as electron acceptor. Pure and mixed cultures were grown in 60 mL serum bottles containing 40 mL reduced, defined mineral salts medium with 5 mM acetate serving as both carbon source and electron donor with an N_2-CO_2 (80:20, vol/vol) headspace (Löffler et al., 1997). Lactate (5 mM) substituted for acetate to grow *G. thiogenes* and consortium BDI. The KB-1 consortium was maintained with methanol as an electron donor, amended to 5 times the electron equivalents required for complete dechlorination. To test the effects of cyanocobalamin (CN-Cbl) on PCE dechlorination, cultures of strain SZ were amended with 0.05 mg mL^{-1} PCE as the sole electron acceptor and 0, 15, and 750 μg CN-Cbl L^{-1}. Chlorinated ethenes were quantified by gas chromatography as described (Amos et al., 2007; Duhamel and Edwards, 2004).

4.5.2 Plasmid stability.

To explore if plasmid pSZ77 was stably maintained (i.e., segregational stability), strain SZ cultures were grown at 35°C with 20 mM sodium fumarate as the electron acceptor in medium with 1,500 μg L^{-1} and without CN-Cbl. When visible turbidity was apparent after 24 to 72 hours, the cultures were consecutively transferred (1% inoculum size, vol/vol) for at least 20 times to fresh medium. Plasmid curing was also attempted by

amending liquid cultures with 0.001-0.1% SDS (w/vol) (El-Mansi et al., 2000) or 0.1 ML-ascorbic acid (Ramesh et al., 2001).

4.5.3 Genomic DNA extraction and PCR.

PCR primers were designed using Primer3 software (http://frodo.wi.mit.edu) based on the predicted open reading frame (ORF) encoding the homolog to the replication initiation protein RepA identified on pSZ77. The primers repA_136F (5'-AGCATCGGTCAGCTGAATCT-3'), and repA_700R (5'-GGTTAGAGCGTGGTGCATTT-3') were used to amplify a 565 bp fragment of the pSZ77 *repA* gene. To test for the presence of *repA* in other cultures, biomass was collected from 2 mL aliquots of pure and mixed cultures by centrifugation at 13,200 rpm for 20 min at room temperature (RT). The pellet was added to a bead tube and the DNA extracted according to the protocol for the PowerSoil DNA isolation kit (Mo Bio Laboratories, Inc.) provided by the manufacturer. PCR reactions were prepared in a volume of 15 µL containing 1x PCR buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 200 nM of each forward and reverse primers, 0.02 U of GoTaq DNA polymerase and 1 µL of DNA template. The PCR thermocycler program for the *repA*-targeted primers was 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 30s, extension at 72°C for 30s, and a final extension at 72°C for 6 min. PCR was also carried out using the primer pair Geo564F/840R (Cummings et al., 2003) that targets the 16S rRNA gene of all *Geobacteraceae*. PCR conditions and the amplification profile used for the Geo564F/840R set were as described for the *repA* primer set.

4.5.4 Plasmid isolation.

For large plasmid isolation using a modified protocol based on the Kieser method (Kieser, 1984; Sobecky et al., 1997), 50 to 200 mL of fumarate-grown SZ culture and 100 mL each of fumarate-grown *G. thiogenes* and the four Fort Lewis *Geobacter* isolates were collected by centrifugation (3,220 x *g*, 30 min). The DNA was isolated as previously described except that the plasmid DNA was precipitated by standard methods (Sambrook and Russell, 2001) and suspended in a final volume of 50–100 µL TE buffer (10 mM Tris, 1 mM EDTA [pH 8.0]). Plasmid DNA extracts were separated by electrophoresis through a 0.7% (w/vol) agarose Tris-acetate gel run in 1x Tris-acetate EDTA (TAE) buffer (Sambrook and Russell, 2001). DNA was visualized by staining in 0.5 µg ethidium bromide per mL TAE buffer solution. Isolate *G. lovleyi* strain KB-1 was grown with 10 mM acetate and 90 mg L^{-1} PCE, and plasmid DNA was extracted from biomass obtained from 1 L of culture suspension using the Qiagen Plasmid Midi Kit and the modified protocol for large inserts (www.qiagen.com/literature/handbooks/default.asp).

4.5.5 Sequencing.

Genomic DNA of *G. lovleyi* strain SZ was extracted from cells grown with acetate and fumarate following established protocols (bacterial genomic DNA isolation using CTAB. http://my.jgi.doe.gov/general/). Sequencing was performed by the Department of Energy's Joint Genome Institute (JGI) using a combination of 454 and Sanger reads with average lengths of 107 and 949 bp, respectively. The total 662,511 reads provided an average 25-fold coverage for the chromosome and 40-fold coverage for the plasmid. The *G. lovleyi* strain SZ genome sequences have been assigned GenBank

accession numbers CP001089 and CP001090 for the chromosome and pSZ77, respectively. DNA was extracted from consortium KB-1 following an established protocol (Wilson, 2001), and clone libraries were generated by the JGI using in-house protocols (www.jgi.doe.gov/sequencing/protocols) and sequenced by the Sanger approach. The strain KB-1 plasmid DNA was incorporated into a bar-coded 454 GS FLX Titanium sequencing run at the Center for Applied Genomics at the University of Toronto. The 454 reads, along with the consortium KB-1 metagenome contigs identified as *Geobacter* plasmid sequences, were aligned against pSZ77 using the Geneious assembly tool (www.geneious.com). Metagenome reads with disagreements with the pSZ77 sequence were verified and removed if necessary. Read depth was at least 3-fold for most of the assembly with a maximum of 12-fold coverage, but as low as single coverage in two regions of less than 136 bp. Amplicons of the 16S rRNA genes of the Ft. Lewis *G. lovleyi* strains Geo7.1, Geo7.2, Geo7.3, and Geo7.4 were sequenced by the Sanger approach and assigned the GenBank accession numbers JN982204 through JN982211.

4.5.6 Computational analyses.

Customized Perl scripts were used to determine G + C percentage of all predicted protein-coding genes on the strain SZ chromosome and pSZ77 with standard deviations from genomic averages computed in R (www.r-project.org/). COGs assignments were determined from the NCBI .ptt files for the *Geobacteraceae* and related non-*Geobacteraceae* chromosomes (AE017180, CP000148, CP000698, CP001390, CP001124, CP001661, CP002479, CP000482, and CP000142) and plasmids (CP000149 CP000483, and CP000484) using customized Perl scripts. Candidate *c*-type cytochromes were determined by searching all strain SZ amino acid

sequences for the CxxCH motif using a customized Perl script, and then further screened for homology to *c*-type cytochromes in the RefSeq database using PSI-Blast (e-value = 1e-11, h = 1e-5) (Altschul et al., 1997) and the PROSITE profile for *c*-type cytochromes (ca.expasy.org/tools/scanprosite/). CxxCH-containing sequences lacking over 50% of PSI-Blast matches annotated as *c*-type cytochromes or lacking recognizable *c*-type cytochrome PROSITE profiles (i.e., PS51007, PS51008, PS51009, and PS51010) were eliminated from further analysis. Trans-membrane regions of predicted outer membrane receptor proteins were determined using Pred-TMBB (Bagos et al., 2004). The nucleotide sequence of pSZ77 and the chromosomal genomic islands on the SZ chromosome were analyzed for repeats using REPuter (Kurtz et al., 2001). Codon adaptation indices (CAI) for strain SZ ORFs encoded on the plasmid and chromosomal genomic islands were computed against a codon usage table based upon all strain SZ chromosomal ORFs using the E-CAI server (Puigbò et al., 2008). Computed CAI were normalized to the expected CAI based upon a 5% level of significance for the bootstrapped set of all ORFs on the SZ chromosome, such that putative foreign genes (i.e. non-*Geobacter*) would score < 1.00 (Puigbò et al., 2008). The origin of replication of pSZ77 was identified using Ori-Finder (Gao et al., 2008), which searches for DnaA-binding sites at regions of GC-skew reversals. The amino acid sequences of replication initiation genes (*repA*) were used to infer plasmid phylogeny and 16S rRNA genes DNA sequences were used to infer genome phylogeny. All alignments were performed in MUSCLE (Edgar, 2004) and used to build trees in Phylip with topology inferred by bootstrapped neighbor-joining and branch lengths computed by maximum likelihood (Felsenstien, 1989). 16S rRNA gene sequences were also used to confirm family and genus affiliation using the Ribosomal

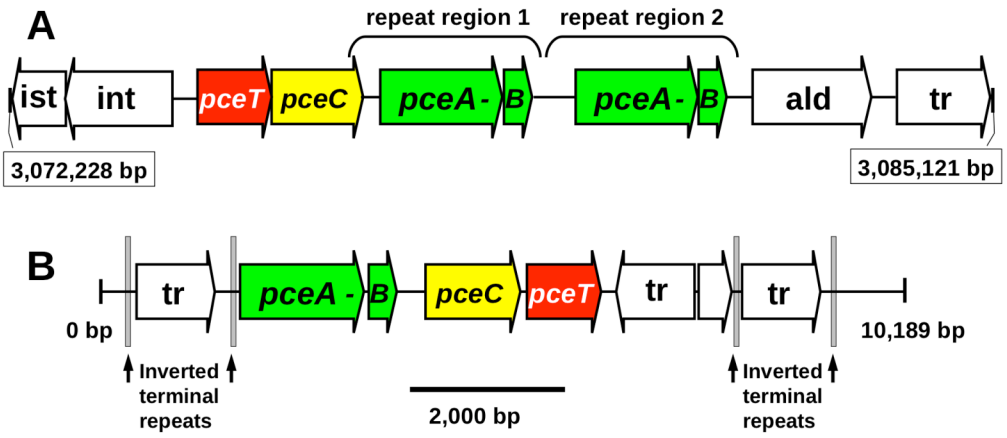
Database Project (Cole et al., 2009). Trees were visualized and formatted using the interactive Tree of Life tool (Letunic and Bork, 2007).

Translations of all predicted GenBank ORFs on pSZ77 were run through BlastP (Altschul et al., 1990) against the nr database with BLOSUM45 and an e-value cutoff of 0.00001. To determine pSZ77 genes and chromosomal genomic islands acquired from *Bacteria* outside of the *Geobacteraceae*, all translated ORFs on pSZ77 and the SZ chromosome were queried using BlastP against the genomes of *Geobacter sulfurreducens* strain PCA (AE017180), *Geobacter metallireducens* strain GS-15 (CP000148), *Geobacter uraniireducens* strain Rf4 (CP000698), *Geobacter* sp. strain FRC-32 (CP001390), *Geobacter bemidjiensis* strain Bem (CP001124), *Geobacter* sp. strain M21 (CP001661), *Geobacter* sp. strain M18 (CP002479), *Pelobacter propionicus* DSM 2379 (CP000482), and *P. carbinolicus* DSM 2380 (CP000142). Pairwise average amino acid identities between predicted proteomes were mined from BlastP output using custom Perl scripts and genome affiliations of top BlastP matches were checked against IMG statistics for *G. lovleyi* strain SZ (img.jgi.doe.gov). Sequence similarities to orthologs for pSZ77 cobalamin biosynthesis genes, pSZ77 *repA*, the chromosomal *pce*-genes, and chromosomal conjugative pilus *tra*-genes were determined using PSI-Blast (Altschul et al., 1997). Further functional inference of amino acid sequences from plasmid- or genomic-island-encoded genes were performed by parsing for conserved domains in PFAM (pfam.janelia.org) and conserved motifs in PROSITE (ca.expasy.org/tools/scanprosite/). A physical map of pSZ77 was plotted using CGview.

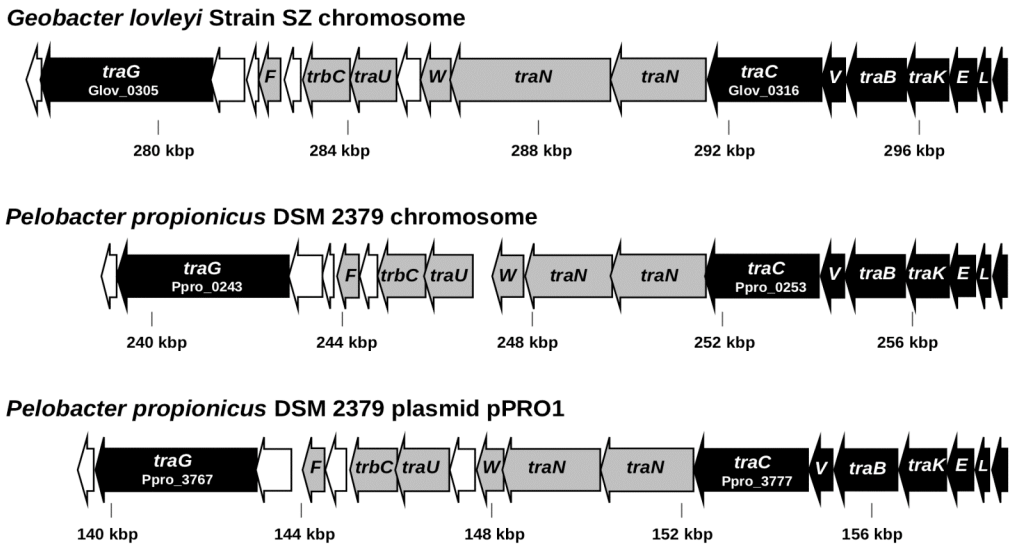
4.6 Chapter 4 Acknowledgements.

We thank the DOE Joint Genome Institute for sequencing the genome of *G. lovleyi* strain SZ. This research was supported by the Strategic Environmental Research and Development Program (SERDP) under contract W912HQ-10-C-0062 (project ER-1586), the U.S. DOE Office of Science (OBER), Subsurface Biogeochemical Research (SBR) Program (grant no. ER64782), and the Government of Canada through Genome Canada and the Ontario Genomics Institute (2009-OGI-ABC-1405).

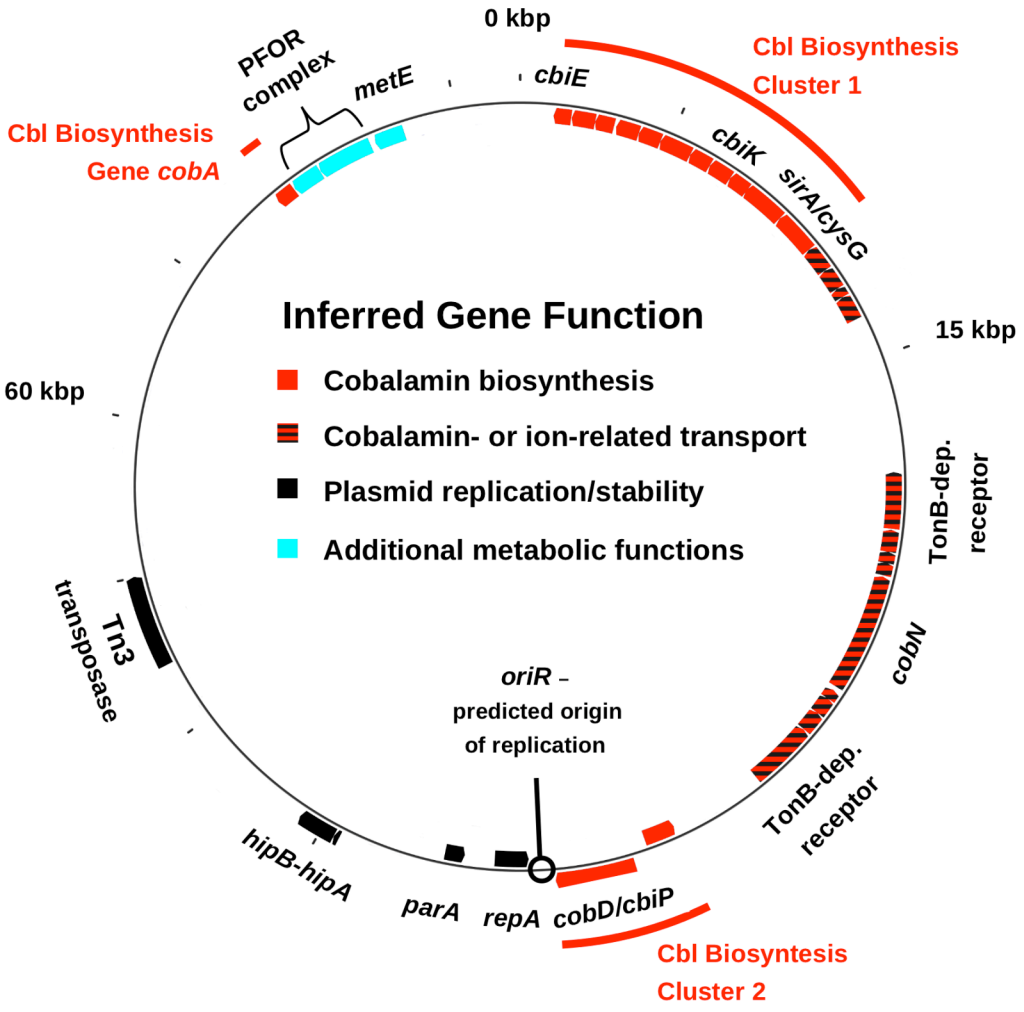
4.7 Chapter 4 Figures.



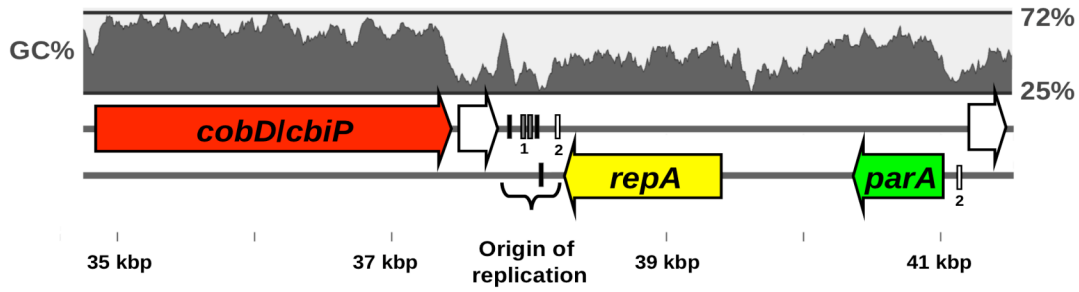
4.7.1 Figure 1. PCE dehalogenase (Pce) genomic island on the *G. lovleyi* strain SZ chromosome. (CP001089; Glov_2866-Glov_2875) (Panel A) compared to the transposon-encoded *pce*-gene cluster from *Desulfitobacterium hafniense* strain TCE1 (AJ439608) (Panel B). The transposase-associated *pce*-gene clusters in the genomes of *D. hafniense* strain Y51 (AP008230) and *Dehalobacter restrictus* PER-K23 (AJ439607) share 99% nucleotide identity with the strain TCE1 *pce*-gene cluster. Inferred gene functions are abbreviated as follows: *pceA* - putative tetrachloroethene (PCE) dehalogenase, *pceB* - membrane anchor protein subunit, *pceT* - peptidylprolyl-*cis*-*trans*-isomerase, *pceC* - FMN-binding and polyferredoxin domain protein, int - integrase, ist - ATPase, tr - transposase, and ald - aldehyde dehydrogenase. The *pceA*-*pceB* ORFs are duplicated within the strain SZ genomic island, and associated with transposase genes, but no inverted repeats were detected. In contrast, the RDase genes in strain TCE1 are associated with both transposases and inverted repeats, and the overall structure of the element in TCE1 resembles a composite transposon.



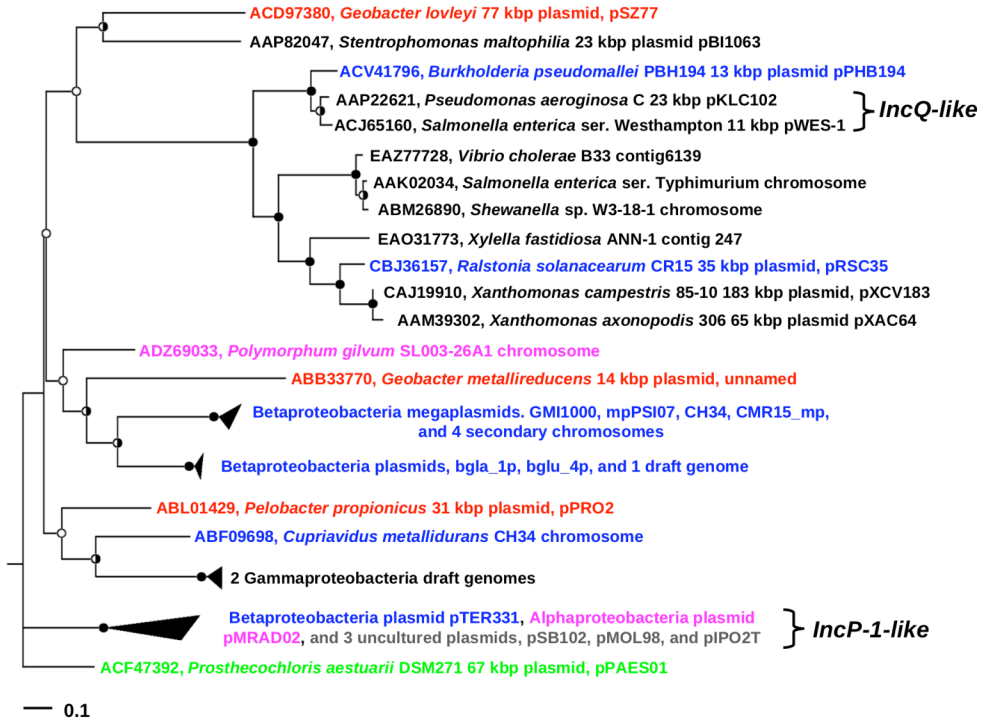
4.7.2 Figure 2. *G. lovleyi* Tra chromosomal genomic island and orthologues. The *G. lovleyi* Strain SZ chromosomal genomic island encoding the putative F-type conjugative pilus *tra*-gene cluster (Glov_0304 to Glov_0322) and the orthologous gene clusters on the chromosome and 202 kbp plasmid pPRO1 of *P. propionicus*. Black ORFs represent genes homologous to conjugative pilus cluster core genes, likely functioning in pore complex assembly and mating pair stabilization. Gray ORFs are homologous to F-type conjugative pilus accessory genes while unshaded ORFs lack inferred function.



4.7.3 Figure 3. Physical map of *G. lovleyi* strain SZ plasmid pSZ77. Cobalamin (Cbl) biosynthesis clusters are red. Transport-related genes such as the TonB-dependent receptors (Glov_3667 and Glov_3675) are red-black barred. Genes likely to play a role in pSZ77 replication or maintenance, such as *repA* (Glov_3681) are solid black. Additional metabolic functions, such as *metE* (cobalamin-independent methionine synthase) and PFOR complex genes (pyruvate ferredoxin/flavodoxin oxidoreductase A and B), are indicated in light blue. Genes lacking predicted function in *Geobacter* metabolism or pSZ77 replicon stability are not shown here, but are included in the pSZ77 map shown in Figure S3.



4.7.4 Figure 4. Predicted replicon region of pSZ77. The DnaA-binding sites found at a low-GC region 3' to the predicted *repA* ORF provide evidence that this region is the plasmid origin of replication (*oriR*). The repeats are not closely spaced, and do not resemble the tandem repeats found in the plasmids of *Geobacter metallireducens* and *Pelobacter propionicus*. Protein coding genes (arrows) are as follows: *repA* – predicted replication initiation protein (yellow), *parA* – predicted partitioning ATPase (green), *cobD/cbiP*: cobalamin synthesis pathway gene (red). The two unshaded genes are annotated as histone-binding proteins. Non-coding regions of interest (rectangles) are as follows: black = predicted DnaA protein-binding site, TT(C|A)TCCAC(A|G), grey = repeat 1, CGTGACAATTCATGC, white = repeat 2, TGTATAGACAATACAATA.



4.7.5 Figure 5. Unrooted proML tree of 35 plasmid replication initiator protein (RepA) sequences. The majority of RepA sequences in the tree are either from the *β-Proteobacteria* (blue, 15 in all) or the *γ-Proteobacteria* (black, 11 in all). Also represented are *α-Proteobacteria* (magenta) and *Chlorobi* (green). The SZ/KB-1 RepA (red) do not cluster with the other two *δ-Proteobacteria* RepA (red) suggesting a possible origin of the pSZ77 replicon region outside of the *δ-Proteobacteria*. Several of the RepA are chromosome-encoded, suggesting origin from mobilizable plasmids such as IncQ. The bootstrap values at each node with >90% support, 50-90% support and <50% support are indicated by black, half-shaded, and white circles, respectively. The scale bar indicates substitutions per site (0.1 = 1 in 10 amino acid residues).

4.8 Chapter 4 Tables.

4.8.1 Table 1. General features of the *Geobacter lovleyi* strain SZ chromosome.

General features	
Size (bp)	3,917,761
G+C percentage	54.8
Total number of predicted protein-coding genes	3,644
Number of pseudogenes	38
Protein-coding genes	
Total intact protein-coding genes	3,606
- matching annotated homologs	3,138
- matching hypotheticals only	296
- with no public database matches	171
- in one or more COGs	2817
- in energy production COGs	279
- in coenzyme metabolism COGs	122
- as predicted *c*-type cytochromes	48
- percent with top BlastP match in *Geobacter* spp.	40%
- percent with top BlastP match in *Pelobacter* spp.	29%
- percent with top BlastP match in other δ-*Proteobacteria*	5%
Non-protein-coding features	
Number of rRNA operons (16S-23S-5S)	2
Number of transfer RNAs	45
CRISPR regions	1

4.8.2 Table 2. General features of the *G. lovleyi* strain SZ plasmid pSZ77.

Size (bp)	77,113
G+C percentage	53.0
Total number of protein-coding genes	81
Number of pseudogenes	2
Protein-coding genes	
Number of intact protein-coding genes	79
- matching annotated homologs	66
- matching genes with hypothetical function only	6
- with no public database matches	7
- in one or more COGs	66
- in energy production COGs	3
- in coenzyme metabolism COGs	18
- predicted as *c*-type cytochromes	0
- percent with top BlastP match in *Geobacter* (7 genomes)	11%
- percent with top BlastP match in *Pelobacter* (2 genomes)	24%
- percent with top BlastP match in other *δ-Proteobacteria*	37%
Non-protein-coding features	
Number of rRNA operons (16S-23S-5S)	0
Number of transfer RNAs	0

4.8.3 Table 3. Predicted energy metabolism genes identified on the genomes of *Geobacter* spp. and *Pelobacter propionicus.* For each genome, number of *c*-type cytochromes with more than 10 predicted heme-binding sites is positively correlated with total predicted c-type cytochromes ($R \approx 0.98$).

Genome	Protein-coding genes in energy production COGs (class C)	Number of predicted *c*-type cytochromes	*c*-type cytochromes with >10 heme-binding motifs
G. lovleyi	279	49	6
G. sulfurreducens	280	89 *	23
G. metallireducens	356	76 *	17
G. uraniireducens	355	104 *	31
P. propionicus	334	19	1

* According to Butler et al. (2010)

4.8.4 Table 4. *G. lovleyi* strain SZ heme and cobalamin biosynthesis. Heme and cobalamin biosynthesis genes on the *G. lovleyi* strain SZ genome in comparison to the model organism for anaerobic cobalamin biosynthesis, *Salmonella enterica* subsp. *enterica* ser. Typhimurium.

	G. lovleyi SZ genome		Homolog in *Salmonella enterica* subsp. *enterica* ser. Typhimurium LT2			
	Chromosome	Plasmid pSZ77	Gene symbol	Function	% Ident. (%Simil.) Chr.[#]	pSZ77
Early heme biosynth.	Glov_0501		*hemD*	Uroporphyrinogen-III synthase	no match	
	Glov_0502		*hemC*	Porphobilinogen deaminase	52(65)	
	Glov_0503		*hemA*	Glutamyl-tRNA reductase	40(62)	
	Glov_0688		*hemB*	δ-Aminolevulinic acid dehydratase	47(63)	
Late heme biosynthesis	Glov_0521		*hemY*	Protoporphyrinogen oxidase	no match	
	Glov_1228		*hemN*	Coproporphyrinogen III oxidase	27(45)	
	Glov_2930		*hemH*	Ferrochelatase	31(47)	
	Glov_3245		*hemL*	Uroporphyrinogen decarboxylase	52(67)	
Cobalamin biosynthesis	Glov_3417	Glov_3646	*cbiE*	Precorrin-6y C5,15-methyl-transferase	32(54)	35(57)
		Glov_3647	*cbiJ*	Precorrin-6x reductase		30(44)
		Glov_3648	*cbiH*	Precorrin-3B C17-methyltransferase		46(63)
		Glov_3649	*cbiG*	Cobalamin biosynthesis protein		31(49)
		Glov_3650	*cbiF*	Precorrin-4 C11-methyltransferase		50(68)
		Glov_3651	*cbiD*	Cobalamin biosynthesis protein		35(51)
		Glov_3652	*cbiL*	Precorrin-2 C20-methyltransferase		33(52)
		Glov_3653	*cbiK*	Anaerobic cobalt chelatase		32(45)
		Glov_3654	*cbiC*	Precorrin-8X methylmutase		34(52)
		Glov_3655	*cbiA*	Cobyrinic acid a,c-diamide synthase		39(55)
		Glov_3656	*sirA/cysG*	Siroheme synthase/ uroporphyrin-III C-methyltransferase	41(57)/40(57)	
	Glov_2767	Glov_3657	*cbiO*	Cobalt ABC transporter, ATPase subunit	38(54) 35(56)	
	Glov_2768	Glov_3658	*cbiQ*	Cobalt ABC transporter, inner membrane sub.	26(44) 23(45)	
		Glov_3659	*cbiN*	Cobalt transport protein		42(57)
	Glov_2769	Glov_3660	*cbiM*	Cobalt transport protein	35(48)	54(68)
	Glov_0692	Glov_3718	*cobA*	Corrinoid:ATP adenosyltransferase	36(54)	46(67)
	Glov_3081		*cobU*	adenosylcobinamide kinase /adenosylcobinamide-phosphate guanylyltransferase	45(57)	
	Glov_3082*	Glov_3678	*cobT*	Nicotinate-nucleotide-DMBA-phosphoribosyltransferase	38(57)	39(57)

Table 4. (continued)

	G. lovleyi SZ genome		Homolog in *Salmonella enterica* subsp. *enterica* ser. Typhimurium LT2			
	Chromosome	Plasmid pSZ77	Gene symbol	Function	% Ident.(%Simil.) Chr.	pSZ77
Cobalamin biosynthesis	Glov_3083		*cobS*	Cobalamin synthase	35(51)	
	Glov_3084		*cobC*	α-Ribazole-5'-phosphate phosphatase	25(43)	
	Glov_3417		*cbiT*	Precorrin decarbocylase		33(54)
	Glov_3553		*cbiB*	Cobalamin biosynthesis protein	46(62)	
	Glov_3554*	Glov_3679	*cobD/cbiP*	L-threonine O-3-phosphate decarboxylase/cobyric acid synthase	34(51)/45(63)	45(62)/35(51)

[#] Chr. = chromosome

* SZ chromosomal *cobT* and *cobD*/*cbiP* share 78% and 77% identity with their respective pSZ77 paralogs

4.9 References.

Aklujkar, M., Krushkal, J., DiBartolo, G., Lapidus, A., Land, M.L., and Lovley, D.R. (2009) The genome sequence of *Geobacter metallireducens*: features of metabolism, physiology, and regulation common and dissimilar to *Geobacter sulfurreducens*. *BMC Microbiology* **9:**109.

Aklujkar, M., Young, N.D., Holmes, D., Chavan, M., Risso, C., Kiss, H.E., Han, C.S., Land, M.L., and Lovley, D.R. (2010) The genome of *Geobacter bemidjiensis*, exemplar for the subsurface clade of *Geobacter* species that predominate in Fe(III)-reducing subsurface environments. *BMC Genomics*, **11:**490.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215:**403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25:**3389–3402.

Amos, B.K., Christ, J.A., Abriola, L.M., Pennell, K.D., and Löffler, F.E. (2007) Experimental evaluation and mathematical modeling of microbially enhanced tetrachloroethene (PCE) dissolution. *Environmental Science Technology* **43:**1977–1985.

Amos, B.K., Suchomel, E.J., Pennell, K.D., Löffler, F.E. (2009) Spatial and temporal distributions of *Geobacter lovleyi* and *Dehalococcoides* spp. during bioenhanced PCE-NAPL dissolution. *Environmental Science Technology* **43:**1977–1985.

Amos, B.K., Sung, Y., Fletcher, K.E., Gentry, T.J., Wu, W-M., Criddle, C.S., Zhou, J., Löffer, F.E. (2007) Detection and quantification of *Geobacter lovleyi* strain SZ: Implications for bioremediation at tetrachloroethene- and uranium-impacted sites. *Applied and Environmental Microbiology* **73:**6898–6904.

Anderson, R.T., Vrionis, H.A., Ortiz-Bernad, I., Resch, C.T., Long, P.E., Dayvaul, R., Karp, K., Marutzky, S., Metzler, D.R., Peacock, A., *et al* (2003) Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer. *Applied and Environmental Microbiology*, **69:**5884–5891.

Bagos, P.G., Liakopoulos, T.D., Spyropoulos, I.C., and Hamodrakas, S.J. (2004) A Hidden Markov Model method, capable of predicting and discriminating ß-barrel outer membrane proteins. *BMC Bioinformatics* **5:**29.

Braun, V. (1995) Energy-coupled transport and signal transduction through the Gram-negative outer membrane via TonB-ExbB-ExbD-dependent receptor proteins. *FEMS Microbiology Reviews* **16:**295–307.

Butler, J.E., Young, N.D., and Lovley, D.R. (2009) Evolution from a respiratory ancestor to fill syntrophic and fermentative niches: Comparative genomics of six *Geobacteraceae* species. *BMC Genomics* **10:**103.

Butler, J.E., Young, N.D., and Lovley, D.R. (2010) Evolution of electron transfer out of the cell: comparative genomics of six *Geobacter* genomes. *BMC Genomics* **11:**40.

Coates, J.D., Phillips, E.J.P., Lonergan, D.J., Jenter, H., and Lovley, D.R. (1996) Isolation of *Geobacter* species from diverse sedimentary environments. *Applied and Environmental Microbiology* **62:**1531–1536.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., *et al* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37:**D141–D145.

Cologgi, D.L., Lampa-Pastrick, S., Speers, A.M., Kelly, S.D., and Reguera, G. (2011) Extracellular reduction of uranium via *Geobacter* conductive pili as a protective cellular mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 15248-15252.

Costanza, J., Fletcher, K.E., Löffler, F.E., and Pennell, K.D. (2009) Fate of TCE in heated Fort Lewis soil. *Environmental Science Technology* **43:**909–914.

Cummings, D.E., Snoeyenbos-West, O.L., Newby, D.T., Niggemyer, A.M., Lovley, D.R., Achenbach, L.A., and Rosenzweig, R.F. (2003) Diversity of *Geobacteraceae* species inhabiting metal-polluted freshwater lake sediments ascertained by 16S rDNA analyses. *Microbial Ecology* **46:**257–269.

Das, A., Silaghi-Dumitrescu, R., Ljungdahl, L.G., and Kurtz, D.M. (2005) Cytochrome bd oxidase, oxidative stress, and dioxygen tolerance of the strictly anaerobic bacterium *Moorella thermoacetica*. *Journal of Bacteriology* **187:**2297–2301.

Debussche, L., Couder, M., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1992) Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a, c-diamide during coenzyme B12 biosynthesis in *Pseudomonas denitrificans*. *Journal of Bacteriology* **174:**7445–7451.

del Solar, G., Giraldo, R., Ruiz-Echevarría, M.J., Espinosa, M., and Díaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews* **62:**434–464.

DeWever, H., Cole, J.R., Fettig, M.R., Hogan, D.A., and Tiedje, J.M. (2000) Reductive dehalogenation of trichloroacetic acid by *Trichlorobacter thiogenes* gen. nov., sp. nov. *Applied and Environmental Microbiology* **66:**2297–2301.

Doublet, B., Granier, S.A., Robin, F., Bonnet, R., Fabre, L., Brisabois, A., Cloeckaert, A., and Weill, F-X. (2009) Novel plasmid-encoded ceftazidime-hydrolyzing CTX-M-53 extended-spectrum β-lactamase from *Salmonella enteric* serotypes Westhampton and Senftenberg. *Antimicrobial Agents and Chemotherapy* **53:**1944–1951.

Duhamel, M. and Edwards, E.A. (2006) Microbial composition of chlorinated ethene-degrading cultures dominated by *Dehalococcoides*. *FEMS Microbiology Ecology*, **58:**538–549.

Duhamel, M., Mo, K., and Edwards, E.A. (2004) Characterization of a highly enriched *Dehalococcoides*-containing culture that grows on vinyl chloride and trichloroethene. *Applied and Environmental Microbiology* **70:**5538–5545.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5:**113.

El-Mansi, M., Anderson, K.J., Inche, C.A., Knowles, L.K., and Platt, D.J. (2000) Isolation and curing of the *Klebsiella pneumonia* large indigenous plasmid using sodium dodecyl sulphate. *Research in Microbiology* **151:**201–208.

El-Naggar, M.Y., Wanger, G., Leung, K.M., Yuzvinsky, T.D., Southam, G., Yang, J., Lau, W.M., Nealson, K.H., and Gorby, Y.A. (2010) Electrical transport along bacterial nanowires from *Shewanella oneidensis* MR-1. *Proceedings of the National Academy of Sciences of the United States of America* **107:**18127–18131.

Felsenstien, J. (1989) PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics* **5:**164–166.

Ferguson, A.D. and Deisenhofer, J. (2002) TonB-dependent receptors–structural perspectives. *Biochimica et Biophysica Acta* **1565:**318–332.

Fletcher, K.E., Costanza, J., Pennell, K.P., and Löffler, F.E. (2011) Electron donor availability for microbial reductive processes following thermal treatment. *Water Research* **45**: 6625-6636.

Furukawa, K., Suyama, A., Tsuboi, Y., Futagami, T., and Goto, M. (2005) Biochemical and molecular characterization of a tetrachloroethene dechlorinating *Desulfitobacterium* sp. strain Y51: a review. *Journal of Industrial Microbiological Technology* **32:**534–541.

Futagami, T., Yamaguchi, T., Nakayama, S-i., Goto, M., and Furukawa, K. (2006) Effects of chloromethanes on growth of and deletion of the *pce* gene cluster in dehalorespiring *Desulfitobacterium hafniense* Strain Y51. *Applied and Environmental Microbiology*, **72:**5998–6003.

Gao, F. and Zhang, C-T. (2008) Ori-Finder: A web-based system for finding *oriC*s in unannotated bacterial genomes. *BMC Bioinformatics*, **9:**79.

Ghai, R., Hain, T., and Chakraborty, T. (2004) GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* **5**:198–203.

Harrison, P.W., Lower, R.P.J., Kim, N.K.D., and Young, J.P.W. (2010) Introducing the bacterial 'chromid' not a chromosome, not a plasmid. *Trends in Microbiology* **18:**141–148.

Hayes, F. (2000) The partition system of multidrug resistance plasmid TP228 includes a novel protein that epitomized an evolutionarily distinct subgroup of the ParA superfamily. *Molecular Microbiology* **37:**528–541.

He, J., Holmes, V.F., Lee, P.K.H., and Alvarez-Cohen, L. (2007) Influence of Vitamin B_{12} and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Applied and Environmental Microbiology* **73:**2847–2853.

Higgs, P.I., Myers, P.S., and Postle, K. (1998) Interactions in the TonB-dependent energy transduction complex: ExbB and ExbD form homomultimers. *Journal of Bacteriology*, **180:**6031–6038.

Inoue, K., Leang, C., Franks, A.E., Woodward, T.L., Nevin, K.P., and Lovley, D.R. (2011) Specific localization of the *c*-type cytochrome OmcZ at the anode surface in current-producing biofilms of *Geobacter sulfurreducens*. *Environmental Microbiology Reports* **3:**211–217.

Kieser, T. (1984) Factors affecting the isolation of cccDNA from *Streptococcus lividans* and *Escherichia coli*. *Plasmid* **12:**19–36.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye. J., and Giegerich, R. (2001) REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* **29:**4633–4642.

Lawley, T.D., Klimke, W.A., Gubbins, M.J., and Frost, L.S. (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters* **224:**1–15.

Lawley, T.D., Gilmour, M.W., Gunton, J.E., Tracz, D.M., and Taylor, D.E. (2003) Functional and mutational analysis of conjugative transfer region 2 (Tra2) from the IncHI1 Plasmid R27. *Journal of Bacteriology* **185:**581–591.

Lawrence, J.G. and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences of the United States of America* **95:**9413–9417.

Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:**127–128.

Löffler, F.E., Champine, J.E., Ritalahti, K.M., Sprague, S.J., and Tiedje, J.M. (1997) Complete reductive dechlorination of 1,2-dichloropropane by anaerobic bacteria. *Applied and Environmental Microbiology* **63:**2870–2875.

Löffler, F.E., Sanford, R.A., and Ritalahti, K.M. (2005) Enrichment, cultivation, and detection of reductively dechlorinating bacteria. *Methods in Enzymology* **397:**77–111.

Lovley, D.R. (2008) Extracellular electron transfer: wires, capacitors, iron lungs, and more. *Geobiology* **6:**225–231.

Mahadevan, R., Bond, D.R., Butler, J.E., Esteve-Nuñez, A., Coppi, M.V., Palsson, B.O., Schilling, C.H., and Lovley, D.R. (2006) Characterization of metabolism in the Fe(III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. *Applied and Environmental Microbiology* **72:**1558–1568.

Maillard, J., Regeard, C., and Holliger, C. (2005) Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1. *Environmental Microbiology* **7:**107–117.

Maillard, J., Schumacher, W., Vazquez, F., Regeard, C., Hagen, W.R., Holliger, C. (2003) Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus*. *Applied and Environmental Microbiology* **69:**4628–4638.

Marietou, A., Richardson, D., Cole, J., and Mohan, S. (2005) Nitrate reduction by *Desulfovibrio desulfuricans*: a periplasmic nitrate reductase system that lacks NapB, but includes a unique tetraheme *c*-type cytochrome, NapM. *FEMS Microbiology Letters* **248:**217–225.

McMurdie, P.J., Behrens, S.F., Holmes, S., and Spormann, A.M. (2007) Unusual codon bias in vinyl chloride reductase genes of *Dehalococcoides* species. *Applied and Environmental Microbiology* **73:**2744–2747.

McMurdie, P.J., Behrens, S., Müller, J.A., Göke, J., Ritalahti, K.M., Wagner, R., Goltsman, E., Lapidus, A., Holmes, S., Löffler, F.E., *et al*: (2009) Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genetics*, **5:**e1000714.

McMurdie, P.J., Hug, L., Edwards, E.A., Holmes, S., and Spormann, A.M. (2011) Site-specific mobilization of vinyl chloride respiration islands by a mechanism common in *Dehalococcoides*. *BMC Genomics* **12:**287.

Mehta, T., Coppi, M,V., Childers, S.E., and Lovley, D.R. (2005) Outer membrane *c*-type cytochromes required for Fe(III) and Mn(IV) oxide reduction in *Geobacter sulfurreducens*. *Environmental Microbiology* **71:**8634–8641.

Mela, F., Fritsche, K., Boersma, H., Elsas, J.Dv., Bartels, D., Meyer, F., Boer, Wd., Veen, J.Av., and Leveau, J.H.J. (2008) Comparative genomics of the pIPO2/pSB102 family of environmental plasmids: sequence, evolution, and ecology of pTer331 isolated from *Collimonas fungivorans* Ter331. *FEMS Microbiology Ecology* **66:**45–62.

Merkl, R. (2006) A comparative categorization of protein function encoded in Bacterial or Archaeal genomic islands. *Journal of Molecular Evolution* **62:**1–14.

Methé, B.A., Nelson, K.E., Eisen, J.A., Paulsen, I.T., Nelson, W., Heidelberg, J.F., Wu, D., Wu, M., Ward, N., Beanan, M.J., *et al* (2003) Genome of *Geobacter sulfurreducens*: Metal reduction in subsurface environments. *Science* **302:**1967–1969.

Morita, Y., Futagami, T., Goto, M., and Furukawa, K. (2009) Functional characterization of the trigger factor protein PceT of tetrachloroethene-dechlorinating *Desulfitobacterium hafniense* Y51. *Applied Microbiology and Biotechnology* **83:**775–781.

Mouser, P.J., Holmes, D.E., Perpetua, L.A., DiDonato, R., Postier, B., Liu, A., and Lovley, D.R. (2009) Quantifying expression of *Geobacter* spp. oxidative stress genes in pure culture and during *in situ* uranium bioremediation. *The ISME Journal* **3:**454–465.

Nordström, K. and Austin, S.J. (1989) Mechanisms that contribute to the stable segregation of plasmids. *Annual Review of Genetics* **23:**37–69.

Nuñez, C., Esteve-Nuñez, A., Giometti, C., Tollaksen, S., Khare, T., Lin, W., Lovley, D.R., and Methé, B.A. (2006) DNA microarray and proteomic analysis of the RpoS regulon in *Geobacter sulfurreducens*. *Journal of Bacteriology* **188:**2792–2800.

Prakash, O., Thomas, G.M., Dalton, D.D., Chin, K-J., Green, S.J., Akob, D.M., Wanger, G., and Kostka, J.E. (2010) *Geobacter daltonii* sp. nov., an Fe(III)- and uranium(VI)-reducing bacterium isolated from a shallow subsurface exposed to mixed heavy metal and hydrocarbon contamination. *International Journal of Systematic and Evolutionary Microbiology* **60:**546–553.

Puigbò, P., Bravo, I.G., and Garcia, S. (2008) E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* **9:**65.

Ramesh, A., Halami, P.M., and Chandrashekar, A. (2001) Ascorbic acid-induced loss of a pediocin-encoding plasmid in *Pediococcus acidilactic* CFR K7. *World Journal of Microbiology and Biotechnology* **16:**695–697.

Reguera, G., McCarthy, K.D., Mehta, T., Nicoll, J.S., Tuominen, M.T., and Lovley, D.R. (2005) Extracellular electron transfer via microbial nanowires. *Nature* **435:**1098–1101.

Sambrook, J. and Russell, D. (2001) Molecular cloning: A laboratory manual., 3rd edn: Cold Spring Harbor Laboratory.

Sanford, R.A., Cole, J.R., and Tiedje, J.M. (2002) Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68:**893–900.

Schaper, S. and Messer, W. (1995) Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *The Journal of Biological Chemistry* **270:**17622–17626.

Scott, A.I. and Roessner, C.A. (2002) Biosynthesis of cobalamin (vitamin B_{12}). *Biochemical Society Transactions*, **30:**613–620.

Sharma, S., Sathyanarayan, B.K., Bird, J.G., Hoskins, J.R., Lee, B. and Wickner, S. (2003) Plasmid P1 RepA is homologous to the F plasmid RepE class of initators. *The Journal of Biological Chemistry* **279:**6027–6034.

Shelobolina, E.S., Vrionis, H.A., Findlay, R.H., and Lovley, D.R. (2008) *Geobacter uraniireducens* sp. nov., isolated from subsurface sediment undergoing uranium bioremediation. *International Journal of Systematic and Evolutionary Microbiology* **58:**1075–1078.

Slater, J.H., Weightman, A.J., and Hall, B.G. (1985) Dehalogenase genes of *Pseudomonas putida* PP3 on chromosomally located transposable elements. *Molecular Biology and Evolution* **2**:557–567.

Sobecky, P.A., Mincer, T.J., Chang, M.C., and Helinski, D.R. (1997) Plasmids isolated from marine sediment microbial communities contain replication and incompatibility regions unrelated to those of known plasmid groups. *Applied and Environmental Microbiology* **63:**888–895.

Stothard, P. and Wishart, D.S. (2005) Circular genome visualization and exploration using CGview. *Bioinformatics* **21:**537-539.

Strycharz, S.M., Woodard, T.L., Johnson, J.P., Nevin, K.P., Sanford, R.A., Löffler, F.E., Lovley, D.R. (2008) Graphite electrode as a sole electron donor for reductive dechlorination of tetrachloroethene by *Geobacter lovleyi*. *Applied and Environmental Microbiology* **74:**5943–5947.

Sung, Y., Fletcher, K.E., Ritalahti, K.M., Apkarian, R.P., Ramos-Hernández, N., Sanford, R.A., Mesba, N.M., and Löffler, F.E. (2006) *Geobacter lovleyi* sp. nov. strain SZ, a novel metal-reducing and tetrachloroethene-dechlorinating bacterium. *Applied and Environmental Microbiology* **69:**2775–2782.

Sung, Y., Ritalahti, K.M., Sanford, R.A., Urbance, J.W., Flynn, S.J., Tiedje, J.M., and Löffler, F.E. (2003) Characterization of two tetrachloroethene-reducing, acetate-oxidizing anaerobic bacteria and their description as *Desulfuromonas michiganensis* sp. nov. *Applied and Environmental Microbiology* **69:**2964–2974.

Thomas, S.H., Sanford, R.A., Amos, B.K., Leigh, M.B., Cardenas, E., and Löffler, F.E. (2010) Unique ecophysiology among U(VI)-reducing bacteria as revealed by evaluation of oxygen metabolism in *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology* **76:**176–183.

Thomas, S.H., Wagner, R.D., Arakaki, A.K., Skolnick, J., Kirby, J.R., Shimkets, L.J., Sanford, R.A., and Löffler, F.E. (2008) The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* **3:**e2103.

Wilson, K. (2001) Preparation of genomic DNA from bacteria. *Molecular Biology*, **2:**4.

CHAPTER 5

COMPARATIVE GENOMICS OF CORRINOID ANABOLISM IN ORGANOHALIDE-RESPIRING *BACTERIA* WITH VARIED ECOPHYSIOLOGIES.

5.1 Abstract.

Anaerobic organohalide respiration is dependent upon the availability of vitamin B_{12} or analogous corrinoids, where some OHRB, such as *Geobacter lovleyi* strain SZ, harbor the *de novo* pathway for corrinoid biosynthesis while others, such as *Dehalococcoides mccartyi* strain 195, are corrinoid auxotrophs. The current study employs genome comparisons, phylogenetic analyses, and a customized comparative method to infer evolutionary history of corrinoid acquisition pathways among all OHRB and their functional relationship to organohalide respiration. Corrinoid biosynthesis and uptake genes appear to have undergone drastic turnover during the evolution of *δ-Proteobacteria* OHRB, where *de novo* corrinoid biosynthesis pathways of *Geobacter lovleyi* strain SZ and *Desulfomonile tiedjei* strain DCB-1 exhibit a history of horizontal gene transfer (HGT). Moreover, *Anaeromyxobacter dehalogenans* strains have lost all corrinoid biosynthesis genes and likely depend upon the corrinoid ABC transport cluster, *btuF-C-D*, for supplying RDases or other corrinoid-dependent genes. *Chloroflexi* OHRB, including *Dehalococcoides mccartyi*, are incapable of *de novo* corrinoid biosynthesis yet harbor a conserved cluster of corrinoid transport and salvage genes, *btuF-C-D-cbiZ-cbiB-cobD-T-S-C-U*. This cluster likely mediates two pathways for corrinoid uptake, *cbiZ-cbiB* horizontally-acquired from *Archaea*, and *cobU* from *Bacteria*. Additional *cbiB* and *cbiZ*, plus *cobA*, are present in multiple copies presumably involved in corrinoid salvage across *Dehalococcoides mccartyi* genomes. Such lateral acquisition and intragenomic duplication of corrinoid salvage genes suggests OHRB utilize extracellular corrinoids from multiple sources. Given that corrinoid salvage/transport genes are widespread and conserved across OHRB, such genes may serve as a means of monitoring or enhancing organohalide degradation activity.

5.2 Introduction.

Organohalide respiring *Bacteria* (OHRB) play essential roles in the anaerobic degradation of anthropogenic organohalide contaminants, such as tetrachloroethene (PCE), trichloroethene (TCE), vinyl chloride (VC) and a diverse range of chlorinated aromatic compounds such as chlorophenols (Smidt and Vos, 2004; Löffler and Edwards, 2006). Reductive dehalogenases (RDases) are the terminal reductase enzymes mediating cleavage of the carbon-halogen bond in organohalide substrates making energy for growth available to the organohalide-respiring cell (Holliger et al., 1999; Smidt and de Vos, 2004). With the exception of the 3-chlorobenzoate RDase from *Desulfomonile tiedjei*, enzymologically-characterized RDase enzymes are dependent upon a bound corrinoid (e.g., cobalamin) for catalytic activity (Ni et al., 1995; Holliger et al., 1999; Banerjee and Ragsdale, 2003). Organohalide respiration has been characterized in genera across four phyla, *Desulfitobacterium* and *Dehalobacter* representing low GC Gram positives (*Firmicutes*), *Sulfurospirillum* from *ε-Proteobacteria*, *Desulfomonile*, *Anaeromyxobacter*, *Geobacter* from *δ-Proteobacteria*, and *Dehalococcoides* and *Dehalogenimonas* from *Chloroflexi* (Shelton and Tiedje, 1984; Cole et al., 1994; Utkin et al., 1994; Scholz-Muramatsu et al., 1995; Maymó-Gatell et al., 1997; Holliger et al., 1998; Sung et al., 2006; Yan et al., 2009). Aside from their phylogenetic diversity, OHRB exhibit varying metabolic characteristics and habitat preferences. The chlorophenol-respiring *δ*-Proteobacterium, *Anaeromyxobacter dehalogenans* 2CP-C, also respires iron, uranium, and sub-atmospheric concentrations of oxygen (Sanford et al., 2002; Wu et al., 2006; Thomas et al., 2008; Thomas et al., 2010). Other OHRB with versatile respiratory capabilities, such as the PCE-respirers, *Geobacter lovleyi* strain SZ and *Desulfitobacterium hafniense* strain Y51, do not respire oxygen, and exhibit a limited

tolerance to oxic conditions (Thomas et al., 2010; Kim et al., 2012). Moreover, *G. lovleyi* strain SZ synthesizes its own corrinoid during organohalide respiration in pure culture (Wagner et al., 2012; Yan et al., 2012). In contrast *Dehalococcoides mccartyi* and other *Chloroflexi* OHRB strains respire solely upon organohalide compounds, are dependent upon extracellular sources of corrinoids, and exhibit extreme sensitivity to oxygen (He et al., 2007; Amos et al., 2008; Löffler et al., 2012; Yan et al., 2012). Hence, OHRB corrinoid metabolism appears to be related to ecology, respiratory metabolism, and possibly, evolutionary history.

Due to the lack of a genetic system for any known OHRB, experimental characterization of genes encoding corrinoid biosynthesis and corrinoid uptake pathways are limited to non-organohalide-respiring model strains. *De novo* corrinoid biosynthesis was first genetically- and enzymologically-characterized as a pathway mediated by 22 genes in "*Pseudomonas denitrificans*" (Blanche et al., 1995). *P. denitrificans* was later found to represent one of two variants among *de novo* cobalamin biosynthesis pathways in which the oxygen-dependent synthesis of the corrin ring is completed prior to insertion of the cobalt center (Scott and Roessner, 2002). The second type of *de novo* cobalamin biosynthesis, characterized in *Salmonella typhimurium*, is defined by insertion of the cobalt center during the early steps of corrin ring biosynthesis (Raux et al., 1999). *S. typhimurium* also harbors transport genes, *btuF* and *btuC-D*, for uptake of extracellular corrinoid sources, including commercial vitamin B_{12} (Bibber et al., 1999). Considering that additional genes, *hemA*, *hemB*, *hemC*, and *hemD*, are required for biosynthesis of the tetrapyrrole precursor to the corrin ring (Panek and O'Brian, 2002), *de novo* corrinoid biosynthesis and corrinoid transport in *S. typhimurium* together involve up to 30 genes. Two genes in *S. typhimurium*, *cobA* and *cobU*, essential to *de novo* corrinoid biosynthesis, also

mediate the uptake (i.e., salvage) of extracellular incomplete corrinoids such cobinamide. The *S. typhimurium* corrinoid adenosyltransferase, CobA, attaches the upper nucleotide loop to the corrin ring as part of both *de novo* biosynthesis and salvage of incomplete corrinoids (Escalante-Semerena et al., 1990). A bifunctional enzyme, CobU, adds the phosphate moiety and guanosine diphosphate to the corrin ring during assembly of the nucleotide loop in both *S. typhimurium* and *P. dentrificans* (Blanche et al., 1991; O'Toole and Escalante-Semerena, 1993). *Archaea* lack the bifuctional *cobU* gene, where the equivalent guanylation step in archaeal *de novo* pathways is mediated by *cobY* (Thomas and Escalante-Semerena, 2000). Accordingly, the archaeal pathway for corrinoid uptake is mediated by CbiZ, cobinamide amidohydrolase, which converts the assimilated corrinoid to cobyric acid, available for incorpration into *de novo* corrinoid biosynthesis via CbiB (Woodson et al., 2003; Woodson and Escalante-Semerena, 2004). While the *cbiZ* gene was initially found only in *Archaea*, cobinamide phosphate synthase, *cbiB*, found in *Bacteria* and *Archaea*, apparently functions in two roles: *de novo* corrinoid biosynthesis and *Archaea*-type corrinoid salvage. Interestingly, an α-Proteobacterium, *Rhodobacter sphaeroides* 2.4.1 harboring the late-cobalt insertion variant of the *de novo* corrinoid biosynthesis and the bifunctional CobU biosynthesis/salvage gene, has an archaeal-type salvage pathway mediated by CbiZ interacting with CbiB (Gray and Escalante-Semerena, 2009a; Gray and Escalante-Semerena, 2009b). In summary, a diverse set of corrinoid metabolism genes have been experimentally-characterized in model *Bacteria* and *Archaea*, serving as a basis for comparison with the genome sequences of OHRB.

As corrinoid-dependent respiration has been described across four distinct bacterial phyla, corrinoid metabolism provides insights into shared genomic and

physiological features of OHRB. A more complete understanding of OHRB corrinoid metabolism may in turn facilitate biochemical or structural characterization of RDase enzymes or enhance organohalide degradation processes at contaminated sites. To this end, the current study employs gene sequence comparisons, database searches, phylogenetic analyses, and a custom surrogate method for inferring evolutionary history of corrinoid metabolism genes. Key questions addressed regarding the evolutionary history of corrinoid biosynthesis or uptake genes among OHRB center on whether such genes exhibit functional or phylogenetic linkage to organohalide respiration and the role played by horizontal gene transfer (HGT). Moreover, these computational approaches represent the first study providing functional prediction for corrinoid biosynthesis or uptake pathways across all sequenced OHRB.

5.3 Results.

5.3.1 Corrinoid biosynthesis and uptake gene repertoire across OHRB genomes.

Differences in genomic repertoire related to corrinoid biosynthesis or extracellular corrinoid uptake suggest variability in strategies for corrinoid acquisition, where all sequenced OHRB of the *Firmicutes* and two OHRB of the *δ-Proteobacteria* are confirmed or predicted to be capable of *de novo* corrinoid biosynthesis (Figure 1). Homologs to all 30 functions required for *de novo* corrinoid biosynthesis or corrinoid transport were found on the genomes of the metabolically-versatile OHRB, *Geobacter lovleyi* SZ (Figure 2 A), *Desulfomonile tiedjei* DCB-1, *Desulfitobacterium dehalogenans*, *D. dichloroeliminans*, *D. metallireducens*, and *D. hafniense* strains Y51 and DCB-2 (Figure 1, Table S1). A complete pathway for corrinoid biosynthesis was also predicted for the obligate OHRB represented in the *Dehalobacter* CF-50

metagenome (Grostern et al., 2010). Homologs to the *Salmonella typhimurium*-type cobalt chelatase (CbiK, (Raux et al., 1997)) were found on the *Geobacter lovleyi* genome and the *Dehalobacter* CF-50 metagenome while the structurally-distinct *Bacillus*-type cobalt chelatase (CbiX, (Leech et al., 2002)) was inferred on the genomes of *Desulfitobacterium* spp. and *Desulfomonile tiedjei*. Moreover, *G. lovleyi*, *D. teidjei*, *Dehalobacter*, and *Desulfitobacterium* spp. harbored homologs to CbiD and CbiG, mediating corrin ring contraction during *de novo* corrinoid biosynthesis (Figure 2 A). Together, homologs to CbiD, CbiG, and the cobalt chelatases CbiK/CbiX indicate these OHRB possess the early cobalt insertion (or anaerobic) variant of the *de novo* corrinoid biosynthesis pathway (Warren et al., 2002; Roessner and Scott, 2006). In summary, the (meta)genomes of *Desulfitobacterium*, *Dehalobacter*, *Geobacter lovleyi*, and *Desulfomonile tiedjei* harbored all functions encoding *de novo* corrinoid biosynthesis, plus the *cobA* and *cobU* genes functioning in both biosynthesis and salvage of extracellular corrinoids.

Although corrinoids are essential to the functionality of RDase complexes implicated in organohalide respiration, all *Chloroflexi* obligate OHRB, plus one group of *δ-Proteobacteria* OHRB, are corrinoid auxotrophs. As described in previous studies, genomes of obligate OHRB from *Dehalococcoides mccartyi* lack the complete set of genes for *de novo* corrinoid biosynthesis (Figure 2 B, (Seshadri et al., 2005; He et al., 2007; Löffler et al., 2012; Yan et al., 2012)) Here, Blast analyses revealed that all *Dehalococcoides mccartyi* strains, along with related *Chloroflexi*, *Dehalogenimonas lykanthroporepellens* strain BL-DC-9 and Candidatus Dehalobium chlorocoercia strain DF-1, possess only partial corrinoid biosynthesis pathways, presumably functioning in the uptake of incomplete, extracellular corrinoids. All seven *Chloroflexi* genomes harbor homologs to genes enabling the addition of the

upper and lower nucleotide loops to corrinoid rings, *btuR*, *cobD*, *cbiB, cobU, cobT, cobC, cobS.* plus the two corrinoid ring amidation genes, *cbiA* and *cbiP* (Figure 2 B, Table S1). Moreover, *Chloroflexi* OHRB were found to harbor *btuF-btuC-btuD*, the corrinoid ABC transporter, *cobU*, functioning in the bacterial-type corrinoid salvage pathway, and *cbiZ*, encoding the cobinamide amidohydrolase mediating a predicted archaeal-type corrinoid salvage pathway (Figure 1) Genes encoding each of these functions were harbored in a 10-gene cluster, *btuF-C-D-cbiZ-cbiB-cobD-T-S-C-U*, conserved across all five *Dehalococcoides mccartyi* and on the draft genome of Candidatus Dehalobium chlorocoercia (Figure 3). These findings provide preliminary evidence that all obligate organohalide-respirers of the *Chloroflexi* are corrinoid auxotrophs while possessing more than one mechanism for uptake of extracellular corrinoid. Genomes of the versatile, microaerophilic chlorophenol-respirers of *Anaeromyxobacter dehalogenans* lacked detectable homologs to any of the genes involved in *de novo* corrinoid biosynthesis yet did possess inferred *btuF-btuC-btuD* corrinoid ABC transport clusters (Figure 1, Table S1). Corrinoid auxotrophy related to the reduced set of genes involved in corrinoid acquisition among *Chloroflexi* OHRB has been confirmed for *Dehalococcoides mccartyi* model strains (Johnson et al., 2009; Yan et al., 2012); However, corrinoid metabolism has not been experimentally-investigated in *Anaeromyxobacter* strains, which appear have undergone a much more drastic reduction in corrinoid acquisition genes.

5.3.2 Horizontal gene transfer of corrinoid biosynthesis and uptake genes.

Genes encoding *de novo* corrinoid biosynthesis on two *δ-Proteobacteria*, the model PCE-respirer, *Geobacter lovleyi* strain SZ, and the chlorobenzoate respirer, *Desulfomonile tiedjei* strain DCB-1, exhibit a history of horizontal gene transfer

(HGT). The corrinoid biosynthesis and transport protein sequences from *G. lovleyi* strain SZ (Figure 2 A, Table 2) shared average identities of 50% and 56% with homologs in *Geobacter sulfurreducens* and *Pelobacter carbinolicus*, respectively (Fig. S1 A). Yet, low average housekeeping protein identities between strain SZ and *P. carbinolicus* (67%) compared with *G. sulfurreducens* (76%), suggest corrinoid metabolism gene evolution among *δ-Proteobacteria* does not match genomic phylogeny (Fig. S1 A). Moreover, 15 of the *G. lovleyi* strain SZ corrinoid biosynthesis genes occur on a 77 kbp plasmid (Wagner et al., 2012) and share top BlastP matches with *Pelobacter carbinolicus*, *Desulfuromonas acetoxidans*, or *Desulfomonile tiedjei*, suggesting lateral acquisition of corrinoid metabolism genes in strain SZ. On the *Desulfomonile teidjei* strain DCB-1 chromosome, 11 out of the 17 total main cluster corrinoid biosynthesis and cobalt-transport genes (Desti_5589-Desti_5606) have their top BlastP matches in *Pelobacter carbinolicus*, *Desulfuromonas acetoxidans*, or *Geobacter lovleyi*. Moreover, nine genes from Desti_5589 to Desti_5606 have three or fewer δ-proteobacterial Blast matches among the top 20 sequences retrieved, suggesting a history of HGT. *D. tiedjei* strain DCB-1 is classed in the *Syntrophaceae* family of the *δ-Proteobacteria*, yet five of its housekeeping proteins have their top BlastP matches among *Geobacter* spp. with an average 66% amino acid identity, possibly a consequence of the small number of available Syntrophaceae genomes (3 at NCBI). Corrinoid salvage or transport genes are found at separate loci on the *Desulfomonile tiedjei* chromosome and exhibit evidence of archaeal origins. For instance, Desti_3359 encodes an iron transport/CbiZ fusion protein (Fec/CbiZ) where both domains share 49% protein identity with Sfum_2743 from *Syntrophobacter fumaroxidans*. CbiZ, the C-terminal region of

Desti_3359 shares 43% protein identity with *Vulcanisaeta distributa* CbiZ and groups with an archaeal cluster of the CbiZ phylogeny (Fig. 3). Yet, deep branching between Desti_3359 CbiZ and its archaeal homolog in *V. distributa* indicates an ancient HGT event. The genes, Desti_5465 to Desti_5466 encode a predicted corrinoid ABC transport complex (BtuF-F-C-D) and have up to 15 of their top 20 BlastP matches in *Archaea*. Hence corrinoid salvage or transport genes of *D. tiedjei* exhibit evidence of HGT from *Archaea*. In summary, genes encoding corrinoid biosynthesis pathways or transport/salvage of *G. lovleyi* strain SZ and *D. tiedjei* strain DCB-1 exhibit a history of HGT although sequencing data are insufficient to determine direction of transfer.

Sequence similarities and phylogenetic trees indicate a subset of corrinoid salvage genes in the obligate organohalide-respirers of the *Chloroflexi* were horizontally-acquired from *Archaea*. All *Dehalococcoides mccartyi* strains shared average protein identities >80% for corrinoid transport/salvage proteins with strain 195 homologs (c.f., >95% for housekeeping proteins, Figure S1 B). The strain 195 corrinoid transport/salvage proteins shared 48%, 50%, and 39% average identities with *Dehalogenimonas lykanthroporepellens*, Candidatus Dehalobium chlorocoercia, and the non-organohalide-respiring *Chloroflexus aurantiacus*, respectively (housekeeping protein avg., 57%-73%). While this decreasing trend in corrinoid protein similarities relative to housekeeping proteins suggests vertical evolution, CbiZ and CbiB from *Chloroflexi* OHRB exhibited apparent phylogenetic relatedness with archaeal and deltaproteobacterial homologs (Figs. 3 and S2). For instance, the CbiZ of *Dehalococcoides mccartyi*, *Dehalogenimonas*, and Dehalobium clustered with the *Archaea*, *Methanocaldococcus* spp., *Methanotorris igneus,* and *Vulcanisaeta distributa* , as well as with the organohalide-respiring δ-Proteobacterium, *Desulfomonile tiedjei* (Fig. 3). CbiZ homologs were lacking in non-OHRB

Chloroflexi, such as *Chloroflexus aurantiacus*, but CbiB of *Dehalococcoides*, *Dehalogenimonas*, and 'Dehalobium' clustered with non-OHRB *Chloroflexi* CbiB (Fig. S2, clade 'Chl'), plus the non-OHRB *δ*-Proteobacterium, *Sorangium cellulosum*. In turn, the entire 'Chl' clade clustered with *Archaea* (Fig. S2). Hence, a subset of the genes comprising the incomplete corrinoid biosynthesis pathway of *Chloroflexi* obligate OHRB appear to have their deepest branches in *Archaea*, presumably the result of ancient HGT.

Evolutionary origins spanning both the *Bacteria* and *Archaea* were inferred for the main corrinoid transport salvage cluster of *Chloroflexi*, *btuF-C-D-cbiZ-cbiB-cobD-T-S-C-U*, based upon a customized log fraction analysis. For each protein sequence represented in the corrinoid transport salvage gene cluster, bit scores of the top 20 archaeal or bacterial BlastP matches were summed and normalized by bit scores of total archaeal or bacterial sequences retrieved. For example, BlastP searching using the *Dehalococcoides mccaryti* strain 195 CbiZ (DET0687) retrieved among its top 20 matches 11 archaeal homologs and 9 bacterial homologs with bit scores totaling 1,664 and 1,027, respectively. Total bit scores for all archaeal and bacterial sequences retrieved by CbiZ, 4,477 for 68 archaeal matches and 9,186 for 147 bacterial matches yielded the fraction: (top 20 archaeal/total archaeal) ÷ (top 20 bacterial/total bacterial) = (1,664 / 4,477) ÷ (1,027 / 9,186). As a result, for all *Chloroflexi* BtuF, BtuD, CbiZ, and CbiB proteins, at least 8 out of the top 20 Blast matches were retrieved from *Archaea*, yielding log-transformed fractions > 0.0 (Fig. 4, red). One or fewer archaeal matches among the top 20 sequences retrieved by BlastP indicated bacterial phylogenetic affiliation, as was found for BtuC, CobT, CobS, CobC, and CobU (Fig. 4, bright green). CobD homologs from *Dehalococcoides mccartyi* strains had three to

four archaeal matches among the top 20, yielding ambiguous phylogenetic affiliation (Fig. 4, brown or dull green). In summary, the log odds fraction distinguished *Chloroflexi* corrinoid transport/salvage genes descended from *Bacteria* from genes in the same cluster exhibiting archaeal origins, validating phylogenetic trees for *Chloroflexi* CbiZ and CbiB (Figs. 3 and S2). Inspection of annotated/curated genes in the IMG database (www.img.jgi.doe.gov; Mavromatis et al., 2009) revealed CbiZ (pfam01955) in 84 out of 111 (76%) archaeal genomes, while only 78 out of 1516 (5%) bacterial genomes harbored CbiZ. Hence, CbiZ would likely be found on bacterial genomes as a result of HGT from *Archaea*. As the CobU homologs of *Chloroflexi* exhibit strictly bacterial evolutionary histories (Fig. S3), the archaeal CbiZ and CbiB in *Dehalococcoides* likely mediate a separate pathway for corrinoid salvage.

Widespread HGT of corrinoid biosynthesis genes was not evident on genomes of Gram positive OHRB, yet the presence of CbiZ in the *Dehalobacter* CF-50 metagenome suggests possible lateral acquisition of the archaeal-type corrinoid salvage. Similarities between corrinoid metabolism proteins versus housekeeping proteins of *Desulfitobacterium hafniense* strain Y51 in comparison to four *Desulfitobacterium* spp. strains, *Dehalobacter* metagenome CF-50, and the non-OHRB *Desulfosporosinus* spp. are consistent with vertical evolution (Figure S1 C). Furthermore, CbiB and CobU from *Desulfitobacterium* cluster with homologs from the related genus, *Desulfosporosinus* (Figs S3 and S4), while *Desulfitobacterium* genomes lack CbiZ (Fig. 1). Yet, the CbiZ homolog from the *Dehalobacter* CF-50 metagenome pairs with the archaeon *Methanococcoides burtonii* in the CbiZ phylogeny (Fig. 3). The log fraction surrogate method associates the *Dehalobacter* CbiZ with *Bacteria* (Fig. 4) despite the three archaeal BlastP matches among their top

20 matches. Hence, HGT of corrinoid salvage genes among Gram positive OHRB is not strongly supported, but cannot be ruled out.

5.3.3 Genomic organization of *Chloroflexi* corrinoid transport and salvage genes.

Comparisons across *Chloroflexi* OHRB genomes revealed varying numbers of duplications in corrinoid salvage genes, often associated with genomic islands or corrinoid-dependent genes, including RDases. Though the main corrinoid transport/salvage cluster on the *D. mccartyi* strain 195 genome is duplicated (DET0650-DET0660 and DET0684-DET0694), the main cluster is associated with a conserved region in all seven *Chloroflexi* genomes harboring the corrinoid-dependent genes, ribonucleotide reductase and carbon monoxide dehydrogenase/acetyl-CoA synthase (Fig. 5, 5 o'clock). At separate loci from the main corrinoid transport/salvage cluster, each of the five *D. mccartyi* genomes harbor duplicated non-identical *cbiZ* and *cbiB* at separate loci from the main transport/salvage cluster. ATP:corrinoid adenosyltransferase genes (*cobA*), absent from the main transport/salvage clusters, are likewise present at multiple loci on *D. mccartyi* genomes. One set of *cobA* and *cbiB* paralogs comprise a two-gene cluster, *cbiB-cobA*, at conserved genome positions across strains 195, CBDB1, BAV1, and VS (Fig. 5, 8 o'clock), while all five *D. mccartyi* genomes harbor an additional, non-operonic *cobA* (Fig. 5, 9 o'clock). Additional *cobA* and *cbiB* homologs (DET0245-DET0246) are associated with one of three non-identical ribonucleotide reductase (*nrd*) duplicates in strains 195, CBDB1, and VS as part of a 13.6 to 15.1 Kbp region (Fig. 5, 2 o'clock), absent in strains BAV1 and GT (Table S3). Multiple cobinamide amidohydrolase genes (*cbiZ*) were located within high-plasticity regions and genomic islands outside

the main corrinoid transport/salvage cluster (Table S3). As a result, *Dehalococcoides* encode variable numbers of *cbiZ*, where strain CBDB1 harbors a total of eight *cbiZ* genes, five of them adjacent to RDase gene loci (Table S3). Similarly, three of the seven *cbiZ* genes on strain 195 were adjacent to RDase genes (Fig. 5, 2, 8, and 11 o'clock). Only three *cbiZ,* two RDase-associated*,* were found on the genomes of strains BAV1 and GT. The detection of multiple *cbiZ*, *cbiB*, and *cobA* on *Dehalococcoides mccartyi* genomes was unexpected in light of their highly reduced genome size (1.34 to 1.47 Mbp, (Löffler et al., 2012)). The conserved genomic positions of the *cobA* and *cbiB*, plus RDase-associated *cbiZ*, suggests possible distinct functions from corrinoid transport or salvage.

The genome of *Dehalogenimonas lykanthroporepellens* strain BL-DC-9 encodes a reduced set of corrinoid biosynthesis genes, several of which are lacking in *D. mccartyi*, suggesting metabolic differences between *Dehalogenimonas* and *Dehalococcoides*. The strain BL-DC-9 corrinoid transport/salvage cluster, *btuF-btuF-C-D-cbiB-cobD-cobT-cobS-cobC-cobU* (Dehly_0718-Dehly_0709), possesses a duplicated corrinoid transport periplasmic component gene, *btuF* (Figure 4, Table S3) while the single *cbiZ* homolog occurs at a distinct locus (Dehly_0902). Strain BL-DC-9 harbors two non-identical ATP:corrinoid adenosyltransferases, one non-operonic, *cobA* (Dehly_1006), and the other downstream of a predicted corrinoid methyltransferase, *cbiE-cobA* (Dehly_0579-Dehly_0580). In contrast with *D. mccartyi* strains, strain BL-DC-9 has genes encoding biosynthesis of uroporhyrinogen III, *hemA-hemC-cysG/hemD-hemB* (Dehly_0453-Dehly_0456) and a precorrin-2 dehydrogenase (Dehly_0452), possibly enabling siroheme biosynthesis. Also present on the BL-DC-9 genome is an incomplete set of heme biosynthesis genes, uroporphyrinogen decarboxylase, *hemL* (Dehly_0457), and coproporphyrinogen

oxidase, *hemN* (Dehly_0576). Moreover, strain BL-DC-9 harbors two putative *c*-type cytochromes, Dehly_0432 and Dehly_0442, with eight and six CxxCH motifs, respectively. By comparison, *D. mccartyi* strains lack confirmed or predicted *c*-type cytochromes. While strain BL-DC-9 harbors a set of corrinoid-dependent genes similar to that found in *D. mccartyi* (Table S4), the heme or siroheme cofactor genes on the strain BL-DC-9 genome may be related to its greater resilience to oxic conditions than observed for *D. mccartyi* strains (Yan et al., 2009; Löffler et al., 2012).

5.4 Discussion.

This study has revealed evolutionary linkage between corrinoid metabolism and organohalide respiration, particularly among OHRB of *δ-Proteobacteria* and *Chloroflexi*, where corrinoid biosynthesis/salvage genes exhibit a history of HGT. Corrinoid biosynthesis/salvage pathways exhibit disparate evolutionary histories across different OHRB phyla; yet, conservation of gene clusters is observed within phyla and corrinoid uptake functions are more or less conserved across all OHRB. For example, the corrinoid transport and salvage gene cluster, *btuF-C-D-*(*cbiZ*)*-cbiB-cobD-T-S-C-U*, was associated with conserved genomic regions across all *Chloroflexi* OHRB (Fig. 4). Moreover, the archaeal-type salvage gene, *cbiZ,* and the bacterial-type *cobU* salvage gene were found across OHRB of three different phyla, the *Chloroflexi* OHRB, the *Dehalobacter* CF-50 metagenome, and *Desulfomonile tiedjei* strain DCB-1. This suggests OHRB often employ multiple pathways for corrinoid uptake, possibly enabling utilization of extracellular corrinoids from more than one source in the microbial community. In particular, the history of HGT from *Archaea* inferred for corrinoid uptake genes, such as *Chloroflexi cbiZ* and *cbiB*, shared and *D.*

tiedjei strain DCB-1 *fec/cbiZ* (Fig. 3) may be related to their ecological association with methanogens and acetogens. The absence of identifiable corrinoid biosynthesis genes in *Anaeromyxobacter dehalogenans* suggests corrinoid anabolism pathways are an unstable trait among some groups of *δ-Proteobacteria*. Yet, the presence of the corrinoid transport genes, *btuF-C-D*, in *Anaeromyxobacter* suggests all OHRB require some type of pathway for corrinoid acquisition (Fig. 1). This suggests a subset of corrinoid salvage or transport genes may serve as biomarkers in monitoring organohalide-respiring populations at contaminated sites. Furthermore, given the absence of a crystal structure for RDase enzymes, increased understanding of OHRB corrinoid metabolism and corrinoid structures associated with RDases may provide needed insights into RDase biochemistry.

Previous studies have shown that the duplicated corrinoid salvage genes of *Dehalococcoides mccartyi* are tightly regulated, while the current study has shown extensive duplications of *cbiZ*, often adjacent to RDase-encoding genes (Figs. 4 and 5). Together, these findings suggest functional linkage between organohalide respiration and corrinoid uptake pathways. The strain 195 *btuF-C-D* transport clusters and the *cbiZ-B* loci in the main corrinoid transport/salvage cluster and *cbiP* (DET0936) were shown to be inhibited when excess (100 μ gL^{-1}) commercial vitamin B_{12} was supplied as the corrinoid source (Johnson et al., 2009). When cell-free mixed culture medium was supplied as a corrinoid source, the bacterial portion of the main transport/salvage cluster, *cobT-cobS-cobC-cobU* was downregulated while *cbiB-cobA* (DET1138-DET1139) was activated (Johnson et al., 2009). By comparison, the entire main corrinoid transport/salvage clusters where inhibited when strain 195 was co-cultured with *Desulfovibrio vulgaris* Hildenborough (Men et al., 2012). Co-culturing

studies have also demonstrated that the *Dehalococcoides* strains BAV1 and FL2 could respire on organohalides using the corrinoids produced by the versatile OHRB, *Geobacter lovleyi* strain SZ, but not from the non-OHRB, *Geobacter sulfurreducens* (Yan et al., 2012). Supplementation of *Dehaloccoccoides* - *G. sulfurreducens* co-cultures with dimethylbenzimidazole restored organohalide respiratory activity, suggesting *Dehalococcoides* RDases may use a corrinoid cofactor structurally similar to commercial vitamin B_{12} (Hodgkin, 1965; Roth et al., 1996; Anderson et al., 2008; Yan et al., 2012). Aside from these recent insights, the function of RDase-associated *cbiZ* are unknown, where 3 out of the 7 strain 195 *cbiZ,* are within two or three reading frames of an RDase gene (Fig. 5). Moreover, the 13 RDase-associated *cbiZ* homologs of *Dehalococcoides mccartyi* cluster together in the phylogenetic tree apart from *cbiZ* associated with the salvage/transport cluster or ribonucletide-reductases (Fig. 3). Given the wide sequence variability among *Dehalococcoides* RDase genes, different RDase enzymes may require a diverse set of modifications for the corrinoid coenzyme. Together, these findings suggest obligate OHRB utilize a range of corrinoids from specific sources within the microbial community, as may occur during *Dehalococcoides* interactions with *Desulfovibrio vulgaris* or *Geobacter lovleyi* (Men et al., 2012; Yan et al., 2012). At the same time, apparent duplications of corrinoid salvage functions on *D. mccartyi* genomes may represent an evolutionary trend towards accommodating a greater diversity of salvaged corrinoids or structural tailoring of corrinoids for a diverse set of RDase enzymes.

Trends in HGT or loss of corrinoid biosynthesis genes among OHRB may serve as an adaptation to the energetic limitations associated with organohalide respiration. Losses of large portions of the *de novo* corrinoid biosynthesis pathway is believed to have occurred among host-associated enteric *γ-Proteobacteria* such as *E.*

coli while *Salmonella enterica* re-acquired *de novo* corrinoid biosynthesis by HGT (Lawrence and Roth, 1996). This study has shown evidence of HGT for corrinoid biosynthesis genes in *Dehalococcoides mccartyi*, *Geobacter lovleyi*, and *Desulfomonile tiedjei*. Horizontal acquisition of the *de novo* corrinoid biosynthesis pathway may enhance OHRB mobility and rapid adaptation to organohalide-contaminated environments, as may be the case for *G. lovleyi* strain SZ (Amos et al., 2009; Wagner et al., 2012; Yan et al., 2012). However, the benefit of a reduced set of corrinoid biosynthesis genes only becomes apparent in light of the energetic costs for *de novo* corrinoid biosynthesis and the ecophysiology of OHRB. Only four corrinoid biosynthesis enzymatic steps require ATP directly, the amidation of cobyrinic acid by CbiA and the amidation of adenosylcobyrinic acid a,c-diamide by CbiP, using two and four ATP molecules, respectively (Scott and Roessner, 2002; Warren et al., 2002). Assembly of the upper nucleotide loop by CobA requires one ATP while initiation of attachment of the lower nucleotide loop by CobU requires 2 ATP (Escalante-Semerena et al., 1990; Scott and Roessner, 2002; Warren et al., 2002). However additional "feeder pathways" to corrinoid biosynthesis require additional ATP molecules. For instance, the formation of the uroporphyrinogen III ring, catalyzed by HemA, HemB, HemC, and HemD, requires at least 8 ATP molecules (Panek and O'Brian, 2002). Furthermore, each methylation reaction from uroporphyrinogen III to cobalt-precorrin 8x requires one S-adenosylmethionine (SAM) coenzyme as carrier of the methyl group. The only known pathway for the regeneration of SAM is the methionine cycle, which uses three ATP molecules to regenerate SAM for subsequent methylations (Loenen, 2006). Given that the corrin ring biosynthesis step catalyzed by CysG performs two methylations, and CbiL, CbiH, CbiF, CbiD, CbiE, and CbiT each perform one methylation, there is a theoretical requirement of 24 ATP molecules to

form the corrin ring. Finally, L-threonine must be phosphorylated to enable CobD to produce 1-amino-2-propanol (Brushaber et al., 1998). The pentose moeity of the lower nucleotide loop is likely derived from the biosynthesis of nicotinate adenine mononucleotide, requiring 3 ATP molecules (Teramoto et al., 2010). Assuming active transport of the cobalt ion, *de novo* biosynthesis of adenosylcobalamin requires a theoretical maximum of 46 ATP. In the absence of acetate, pyruvate or other fermentable electron donors, organohalide respiration represents an inefficient means of energy production, yielding less than one ATP per carbon-halogen bond cleaved (van den Pas et al., 2001; John et al., 2006). Hence, *de novo* corrinoid biosynthesis may be prohibitively costly for obligate OHRBs, with respiration restricted to organohalides as electron acceptors.

While mechanisms for corrinoid uptake are conserved across all OHRB (Fig. 1) and corrinoid uptake genes may be adjacent to RDases (Fig. 5), functional linkage between corrinoid uptake genes and other corrinoid-dependent genes cannot be ruled out. RDases and ribonucleotide reductases (Nrd) are the only corrinoid-dependent genes found across all OHRB genomes. Nonetheless, corrinoid-dependent enzymes involved in carbon metabolism are represented among OHRB of each phylum. For example, carbon monoxide dehydrogenase/acetyl-CoA synthase (AcsD) occurs on the *Desulfomonile tiedjei* genome, all *Firmicutes* OHRB genomes, and all *Chloroflexi* OHRB genomes (Table S4). Furthermore, the main corrinoid salvage/transport clusters of *Dehalococcoides mccartyi* and *Dehalogenimonas lykanthroporepellens* appear to be linked with both Nrd and AcsD (Fig. 5). The metabolically-versatile *Desulfitobacterium* spp. appear to harbor the most diverse set of corrinoid-dependent enzymes implicated in biodegradation processes (Table S4), which include PCE RDases (PceA), chlorophenol RDases (CprA), ethanolamine-ammonia lyase (EutB),

and O-demethylase (OdmA) (Villemur et al., 2006; Kim et al., 2012). Notably, *Desulfitobacterium* spp. in this study exhibited no evidence of corrinoid biosynthesis gene losses nor HGT, suggesting the large set of corrinoid-dependent genes among *Desulfitobacterium* would make *de novo* corrinoid biosynthesis an essential trait. Corrinoid-auxotrophs of the *Chloroflexi* and the *Anaeromyxobacter* genus exhibit both genomic and experimental evidence for metabolic dependence on corrinoids. Given the presence of four predicted corrinoid-dependent proteins on the genomes of *Anaeromyxobacter dehalogenans* strains, NrdJ, MutA, MetH, and the 2-chlorophenol RDases (Table S4), the complete absence of corrinoid biosynthesis genes from the *Anaeromyxobacter* genomes was surprising. In their preferred microaerobic habitats (Thomas et al., 2010), *A. dehalogenans* strains may encounter ample sources of extracellular corrinoids requiring minimal modification. Paradoxically, the obligate organohalide-respiring *Dehalococcoides mccartyi* harbor the largest predicted corrinoid-dependent proteome among all characterized domains of life (Zhang et al., 2009) as a direct consequence of apparent RDase gene duplications and/or horizontal acquisition (Seshadri et al., 2005; McMurdie et al., 2009; McMurdie et al., 2011). Yet, no *Chloroflexi* OHRB isolate exhibits the capacity for *de novo* corrinoid biosynthesis (Löffler et al., 2012). Further experimental and genomics studies will be needed to elucidate the metabolism and evolutionary origins of *Chloroflexi* OHRB, providing insights into the optimal conditions for *in situ* mitigation of organohalide-contaminated sites.

5.5 Methods.

Predicted proteomes of the OHRB, *Geobacter lovleyi* SZ (NC_010814 and NC_010815), *Anaeromyxobacter dehalogenans* strains 2CP-C, 2CP-1, and K (NC_007760, NC_011891, and NC_011145), *Desulfomonile tiedjei* DCB-1

(NC_018025 and NC_018026), *Dehalococcoides mccartyi* (NC_002936, NC_007356, NC_009455, NC_013552, and NC_013890), and *Desulfitobacterium* spp. (NC_007907, NC_011830, NC_018017, NZ_AGJE00000000, and NZ_AGJB00000000), and the non-OHRB, *Geobacter sulfurreducens* PCA (AE017180), *Chloroflexus aurantiacus* (NC_010175), and *Bacillus megaterium* (NC_014103), were retrieved from GenBank. An e-value cutoff of 1e-10 was used in Blast or PSI-Blast queries employed for functional prediction while all matches below the default e-value of 10.0 were retained for purposes of inferring taxonomic affiliation (Altschul et al., 1990; Altschul et al., 1997). Experimentally-characterized cobalamin biosynthesis genes from *Salmonella typhimurium* strain LT2 (AE006468) and curated cobalamin biosynthesis genes from the model OHRB *Geobacter lovleyi* strain SZ and *Dehalococcoides mccartyi* strain 195 were used to query the proteomes of OHRB (Maymó-Gatell et al., 1997; Sung et al., 2006; Aklujkar et al., 2009; Löffler et al., 2012; Wagner et al., 2012). Experimentally-characterized cobalamin salvage pathway genes from *Methanosarcina mazei* Go1 (AE008384), *Halobacterium* sp. NRC-1 (AE004437, AF016485, and AE004438), and *Rhodobacter sphearoides* 2.4.1 (CP000143 and CP000144) were also used to query OHRB proteomes. BlastP identities across *Chloroflexi* genomes were plotted and visualized using GenomeViz (Ghai et al., 2004). To investigate differences in bacterial versus archaeal phylogenetic affiliation for *Chloroflexi* corrinoid transport-salvage gene clusters, a custom, Blast-based surrogate method was developed. Amino acid sequences of the transport-salvage cluster conserved across *Chloroflexi* OHRB, matching DET0650-DET0660 or DET0684-DET0694 on *D. mccartyi* strain 195, were queried against RefSeq using Blast. To compute log fraction phylogenetic affiliation, species names and bit scores

were parsed from Blast output using custom Perl scripts and a log fraction ratio of archaeal versus bacterial affiliation was computed as follows:

Sum of archaeal bit scores / Total bit scores

Sum of bacterial bit scores / Total bit scores

Top twenty = Sum of top twenty non-OHRB Blast matches

weighted bacterial scores = bacterial bit scores in Top Twenty/20 + 0.05

weighted archaeal scores = archaeal bit scores in Top Twenty/20 + 0.05

Fraction of bacterial in Top Twenty = weighted bacterial scores/TS

Fraction of archaeal in Top Twenty = weighted archaeal scores/TS

bacterial Log Fraction = Fraction of bacterial in Top Twenty / Sum of bacterial

archaeal Log Fraction = Fraction of archaeal in Top Twenty / Sum of archaeal

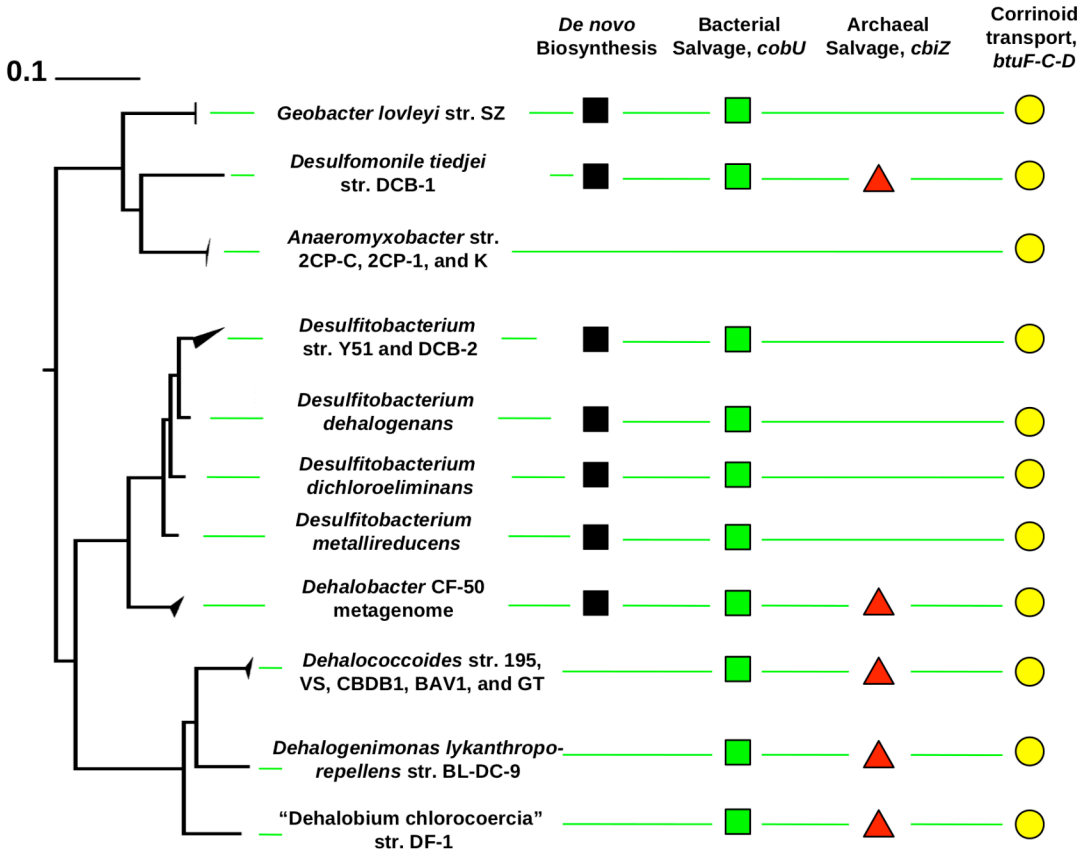
Log Fraction Phylogenetic Affiliation = archaeal LogOdds / bacterial LogOdds

To construct protein and DNA phylogenies, 16S rRNA gene trees and amino acid sequence trees were constructed from in Phylip from Muscle alignments as described (Felsenstein, 1989; Edgar, 2004; Wagner et al., 2012).

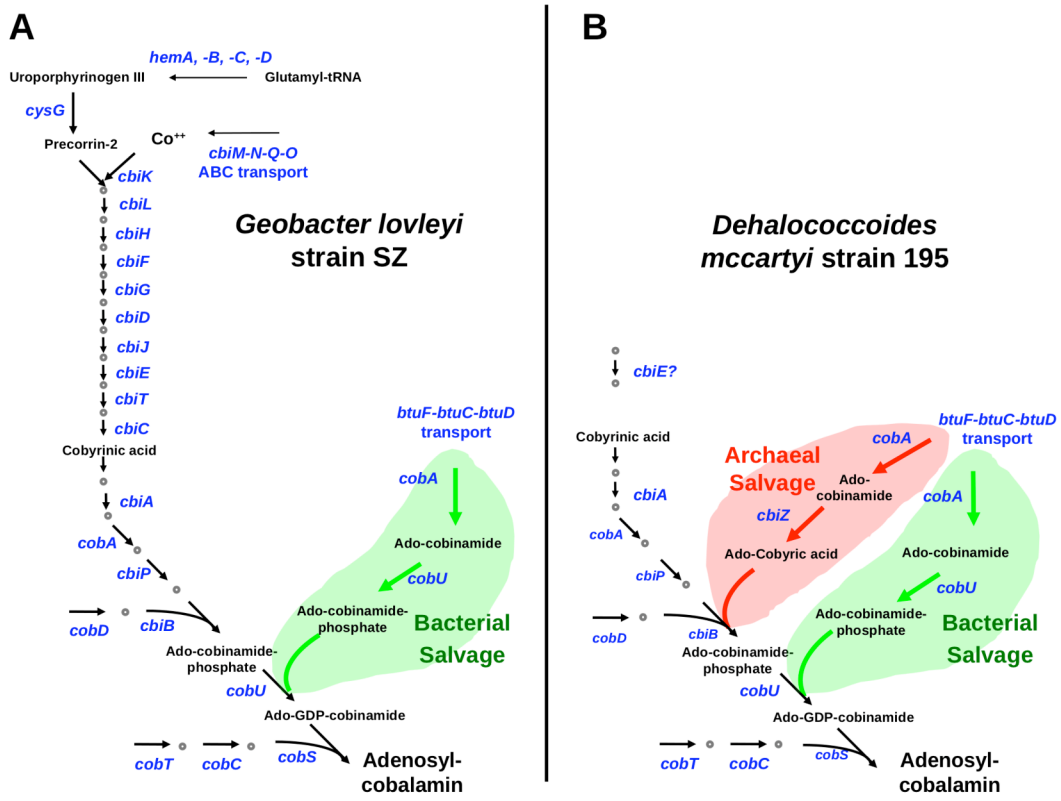
5.6 Chapter 5 Acknowledgements.

This research was supported by the Strategic Environmental Research and Development Program (SERDP) (project ER-1586).

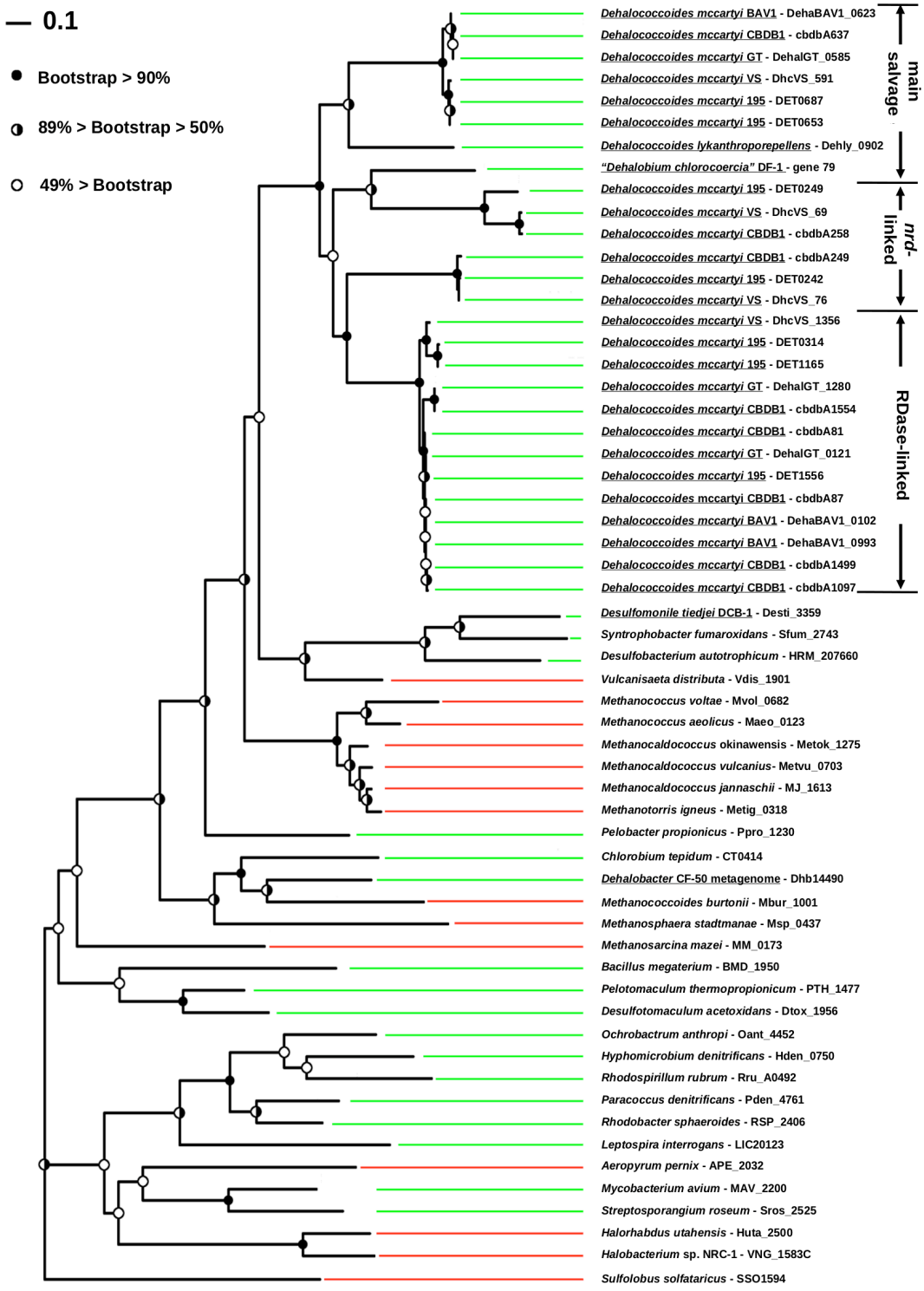
5.7 Chapter 5 Figures.



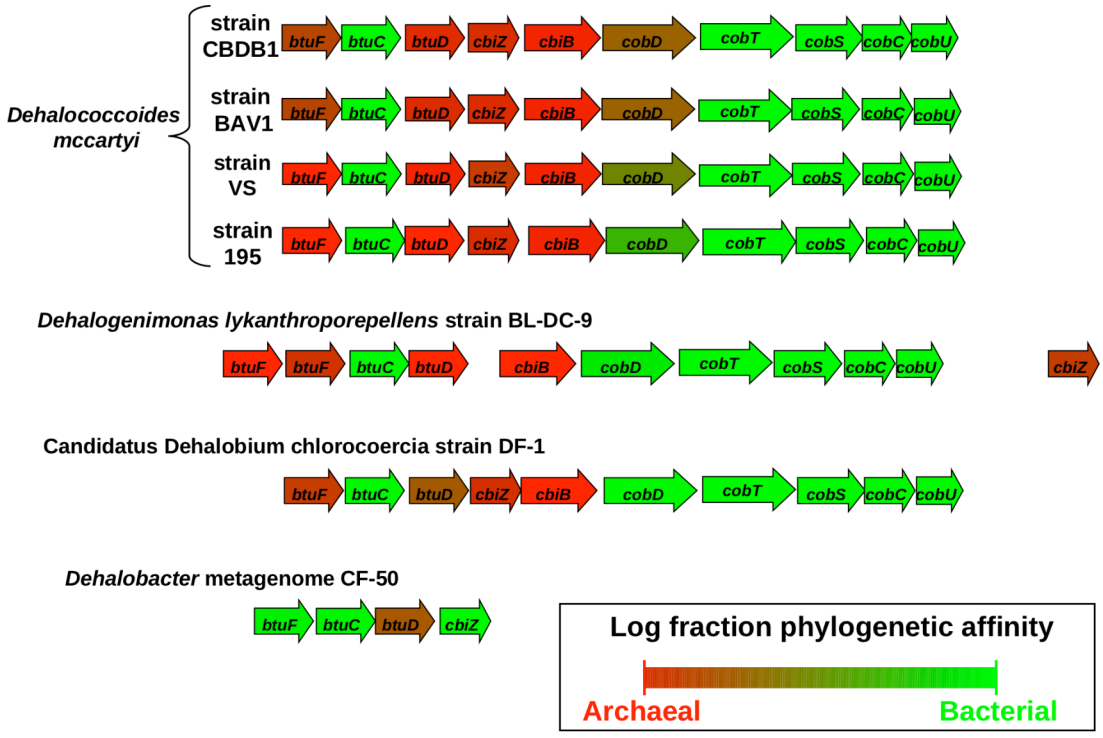
5.7.1 Figure 1. Genomic repertoire of corrinoid biosynthesis and corrinoid salvage/transport genes for organohalide-respiring *Bacteria* (OHRB). The tree at the left margin of the figure table is based upon the 16S rRNA gene phylogeny of the OHRBs shown (0.1 = 100 nucleotide changes per 1000). Black squares next to species/genus name indicates the presence of the full genomic repertoire for *de novo* corrinoid biosynthesis, from delta-aminole-vulinate to adenosylcobalamin, involving up to 30 genes. Green squares indicate the presence of CobU, bifunctional cobinamide kinase/cobinamide phosphate guanylyltransferase (EC:2.7.1.156; EC: 2.7.7.62), mediating the bacterial salvage pathway. Red triangles indicate the presence of a predicted Archaeal-like salvage pathway, mediated by CbiZ, cobinamide amidohydrolase (EC:3.5.1.90).



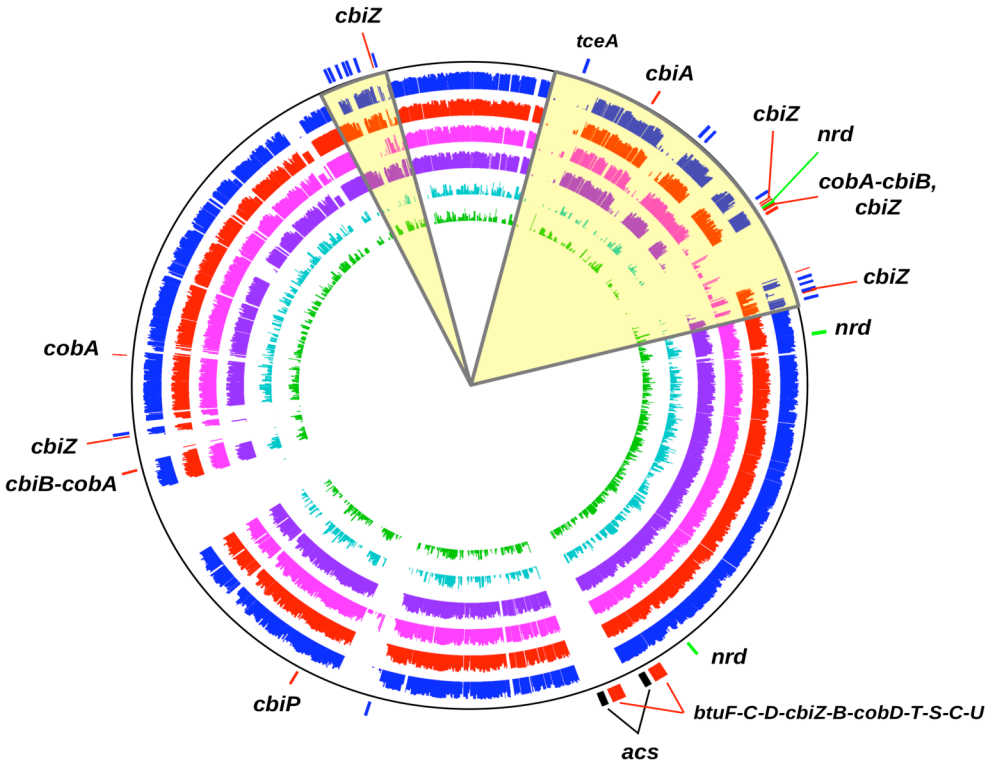
5.7.2 Figure 2. Predicted corrinoid biosynthesis/salvage/transport pathways in model OHRB strains. A. Proteins inferred on genome of *Geobacter lovleyi* strain SZ (NC_010814 and NC_010815) mediating *de novo* corrinoid biosynthesis and the bacterial-type salvage pathway for incomplete corrinoids. **B.** Proteins on *Dehalococcoides mccartyi* strain 195 genome (NC_002936) enabling only an incomplete pathway for corrinoid biosynthesis likely to mediate two separate corrinoid salvage pathways, bacterial- and archaeal-type.


— 0.1
Bootstrap > 90%
89% > Bootstrap > 50%
49% > Bootstrap
Dehalococcoides mccartyi BAV1 - DehaBAV1_0623
Dehalococcoides mccartyi CBDB1 - cbdbA637
Dehalococcoides mccartyi GT - DehalGT_0585
Dehalococcoides mccartyi VS - DhcVS_591
Dehalococcoides mccartyi 195 - DET0687
Dehalococcoides mccartyi 195 - DET0653
Dehalococcoides lykanthroporepellens - Dehly_0902
"Dehalobium chlorocoercia" DF-1 - gene 79
Dehalococcoides mccartyi 195 - DET0249
Dehalococcoides mccartyi VS - DhcVS_69
Dehalococcoides mccartyi CBDB1 - cbdbA258
Dehalococcoides mccartyi CBDB1 - cbdbA249
Dehalococcoides mccartyi 195 - DET0242
Dehalococcoides mccartyi VS - DhcVS_76
Dehalococcoides mccartyi VS - DhcVS_1356
Dehalococcoides mccartyi 195 - DET0314
Dehalococcoides mccartyi 195 - DET1165
Dehalococcoides mccartyi GT - DehalGT_1280
Dehalococcoides mccartyi CBDB1 - cbdbA1554
Dehalococcoides mccartyi CBDB1 - cbdbA81
Dehalococcoides mccartyi GT - DehalGT_0121
Dehalococcoides mccartyi 195 - DET1556
Dehalococcoides mccartyi CBDB1 - cbdbA87
Dehalococcoides mccartyi BAV1 - DehaBAV1_0102
Dehalococcoides mccartyi BAV1 - DehaBAV1_0993
Dehalococcoides mccartyi CBDB1 - cbdbA1499
Dehalococcoides mccartyi CBDB1 - cbdbA1097
Desulfomonile tiedjei DCB-1 - Desti_3359
Syntrophobacter fumaroxidans - Sfum_2743
Desulfobacterium autotrophicum - HRM_207660
Vulcanisaeta distributa - Vdis_1901
Methanococcus voltae - Mvol_0682
Methanococcus aeolicus - Maeo_0123
Methanocaldococcus okinawensis - Metok_1275
Methanocaldococcus vulcanius- Metvu_0703
Methanocaldococcus jannaschii - MJ_1613
Methanotorris igneus - Metig_0318
Pelobacter propionicus - Ppro_1230
Chlorobium tepidum - CT0414
Dehalobacter CF-50 metagenome - Dhb14490
Methanococcoides burtonii - Mbur_1001
Methanosphaera stadtmanae - Msp_0437
Methanosarcina mazei - MM_0173
Bacillus megaterium - BMD_1950
Pelotomaculum thermopropionicum - PTH_1477
Desulfotomaculum acetoxidans - Dtox_1956
Ochrobactrum anthropi - Oant_4452
Hyphomicrobium denitrificans - Hden_0750
Rhodospirillum rubrum - Rru_A0492
Paracoccus denitrificans - Pden_4761
Rhodobacter sphaeroides - RSP_2406
Leptospira interrogans - LIC20123
Aeropyrum pernix - APE_2032
Mycobacterium avium - MAV_2200
Streptosporangium roseum - Sros_2525
Halorhabdus utahensis - Huta_2500
Halobacterium sp. NRC-1 - VNG_1583C
Sulfolobus solfataricus - SSO1594
main salvage
nrd-linked
RDase-linked

5.7.3 Figure 3. Bootstrapped phylogeny of 51 characterized and putative cobinamide amidohydrolases (CbiZ, EC:3.5.1.90). The scale bar indicates change in 10% of aligned amino acid sites. Closed circles indicate at least 90% bootstrap support (out of 100), half-closed circles indicate 50% to 89% bootstrap support, and open circles indicate less than 50% bootstrap support. Note that out of 25 CbiZ homologs encoded on genomes of *Dehalococcoides mccartyi*, 13 are encoded at loci adjacent to RDase genes. OHRB genome/strain names are underlined.



5.7.4 Figure 4. Main corrinoid transport/salvage cluster, *btuF-C-D-cbiZ-cbiB-cobD-cobT-cobS-cobC-cobU*. Computed log fraction relatedness to *Archaea* denoted on red-green color scale where bright red denotes highly probable Archaeal origin (i.e., 10 or more Archaeal matches among top 20 Blast matches). Bright green denotes highly probable bacterial origin, typically fewer than 3 Archaeal matches among top 20 Blast matches. Note displacement of gene encoding cobinamide amidohydrolase, *cbiZ*, in *Dehalogenimonas lykanthroporepellens* strain BL-DC-9.



5.7.5 Figure 5. *Dehalococcoides mccartyi* corrinoid salvage genes and corrinoid-dependent genes. Genome organization of corrinoid transport and salvage genes and corrinoid-dependent genes of *Chloroflexi* using *Dehalococcoides mccartyi* strain 195 as a scaffold. Outermost circle shows strain 195 genomic positions of RDases (blue), ribonucleotide reductases (green), acetyl-CoA synthases (black), and corrinoid transport or salvage genes (red). From outside to inside, second circle (blue) shows strain 195 protein identities to strain VS, third circle (red) shows identities to strain CBDB1, fourth circle (magenta) shows BAV1, fifth circle (violet) shows strain GT, sixth circle (cyan) shows *Dehalogenimonas lykanthroporepellens* BL-DC-9, and the innermost circle shows Candidatus Dehalobium chlorocoercia strain DF-1. Highlighted arcs from 1 o'clock to 2 o'clock and at 11 o'clock indicate high plasticty regions which harbor the majority of RDases in *Dehalococcoides mccartyi* genomes. Note that only strain 195 has the duplicated main corrinoid transport/salvage cluster (5 o'clock).

5.8 References.

Aklujkar, M., Krushkal J., DiBartolo G., Lapidus A., Land M. L. and Lovley D. R. (2009). The genome sequence of *Geobacter metallireducens*: features of metabolism, physiology, and regulation common and dissimilar to *Geobacter sulfurreducens*. *BMC Microbiology* **9**: 109.

Altschul, S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Altschul, S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Amos, B. K., Ritalahti K. M., Cruz-Garcia C., Padilla-Crespo E. and Löffler F. E. (2008). Oxygen effect on *Dehalococcoides* viability and biomarker quantification. *Environmental Science and Technology* **42**: 18-26.

Amos, B. K., Suchomel E. J., Pennell K. D. and Löffler F. E. (2009). Spatial and temporal distributions of *Geobacter lovleyi* and *Dehalococcoides* spp. during bioenhanced PCE-NAPL dissolution. *Environmental Science Technology* **43**: 1977-1985.

Anderson, P. J., Lango J., Carkeet C., Britten A., Kräutler B., Hammock B. D. and Roth J. R. (2008). One pathway can incorporate either adenine or dimethylbenzimidazole as an α-axial ligand of B_{12} cofactors in *Salmonella enterica*. *Journal of Bacteriology* **190**: 1160-1171.

Banerjee, R. and Ragsdale S. W. (2003). The many faces of vitamin B_{12}: Catalysis by cobalamin-dependent enzymes. *Annual Reviews of Biochemistry* **72**: 209-247.

Bibber, M. v., Bradbeer C., Clark N. and Roth J. R. (1999). A new class of cobalamin transport mutants (*btuF*) provides evidence for a periplasmic binding protien in *Salmonella typhimurium*. *Journal of Bacteriology* **181**: 5539-5541.

Blanche, F., Cameron B., Crouzet J., Debussche L., Thibaut D., Vuilhorgne M., Leeper F. J. and Battersby A. R. (1995). Vitamin B_{12}: how the problem of its biosynthesis was solved. *Angewandte Chemie International Edition in English* **34**: 383-411.

Blanche, F., Debussche L., Famechon A., Thibaut D., Cameron B. and Crouzet J. (1991). A bifunctional protein from *Pseudomonas denitrificans* carries cobinamide kinase and cobinamide phosphate guanylyltransferase activities. *Journal of Bacteriology* **173**: 6052-6057.

Brushaber, K. R., O'Toole G. A. and Escalante-Semerena J. C. (1998). CobD, a novel enzyme with L-threonine-O-3-phosphate decarboxylase activity, is responsible for the synthesis of (R)-1-amino-2-propanol O-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. *The Journal of Biological Chemistry* **273**: 2684-2691.

Cole, J. R., Cascarelli A. L., Mohn W. W. and Tiedje J. M. (1994). Isolation and characterization of a novel bacterium growing via reductive dehalogenatioin of 2-chlorophenol. *Applied and Environmental Microbiology* **60**: 3536-3542.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Escalante-Semerena, J. C., Suh S.-J. and Roth J. R. (1990). *cobA* function is required for both *de novo* cobalamin biosynthesis and assimilation of exogeneous corrinoids in *Salmonella typhimurium*. *Journal of Bacteriology* **172**: 273-280.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (version 3.2). *Cladistics* **5**: 164-166.

Ghai, R., Hain T. and Chakraborty T. (2004). GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* **5**: 198-203.

Gray, M. J. and Escalante-Semerena J. C. (2009a). The cobinamide amidohydrolase (cobyric acid-forming) CbiZ enzyme: a critical activity of the cobamide remodelling system of *Rhodobacter sphaeroides*. *Molecular Microbiology* **74**: 1198-1210.

Gray, M. J. and Escalante-Semerena J. C. (2009b). In vivo analysis of cobinamide salvaging in *Rhodobacter sphaeroides* 2.4.1. *Journal of Bacteriology* **191**: 3842-3851.

Grostern, A., Duhamel M., Dworatzek S. and Edwards E. A. (2010). Chloroform respiration to dichloromethane by a *Dehalobacter* population. *Environmental Microbiology* **12**: 1053-1060.

He, J., Holmes V. F., Lee P. K. H. and Alvarez-Cohen L. (2007). Influence of Vitamin B_{12} and cocultures on the growth of *Dehalococcoides* isolates in defined medium. *Applied and Environmental Microbiology* **73**: 2847-2853.

Hodgkin, D. C. (1965). The X-ray analysis of complicated molecules. *Science* **150**: 979-988.

Holliger, C., Hahn D., Harmsen H., Ludwig W., Schumacher W., Tindall B., Vazquez F., Weiss N. and Zehnder A. J. B. (1998). *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetrachloroethene and trichloroethene in an anaerobic respiration. *Archives of Microbiology* **169**: 313-321.

Holliger, C., Wohlfarth G. and Diekert G. (1999). Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiology Reviews* **22**: 383-398.

John, M., Schmitz R. P. H., Westermann M., Richter W. and Diekert G. (2006). Growth substrate dependent localization of tetrachloroethene reductive dehalogenase in *Sulfurospirillum multivorans*. *Archives of Microbiology* **186**: 99-106.

Johnson, D. R., Nemir A., Andersen G. L., Zinder S. H. and Alvarez-Cohen L. (2009). Transcriptomic microarray analysis of corrinoid responsive genes in *Dehalococcoides ethenogenes* strain 195. *FEMS Microbiology Letters* **294**: 198-206.

Kim, S.-H., Harzman C., Davis J. K., Hutcheson R., Broderick J. B., Marsh T. L. and Tiedje J. M. (2012). Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiology* **12**: 21.

Lawrence, J. G. and Roth J. R. (1996). Evolution of coenzyme B_{12} biosynthesis among enteric bacteria: Evidence for loss and reacquisition of a multigene complex. *Genetics* **142**: 11-24.

Leech, H. K., Raux-Deery E., Heathcote P. and Warren M. J. (2002). Production of cobalamin and sirohaem in *Bacillus megaterium*: an investigation into the role of the branchpoint chelatases sirohydrochlorin ferrochelatase (SirB) and sirohydrochlorin cobalt chelatase (CbiX). *Biochemical Society Transactions* **30**: 610-613.

Loenen, W. A. M. (2006). S-Adenosylmethionine: jack of all trades and master of everything? *Biochemical Society Transactions* **34**: 330-333.

Löffler, F. E. and Edwards E. A. (2006). Harnessing microbial activities for environmental cleanup. *Current Opinion in Biotechnology* **17**: 274-284.

Löffler, F. E., Yan J., Ritalahti K. M., Adrian L., Edwards E. A., Konstantinidis K. T., et al. (2012). *Dehalococcoides mccartyi* gen. nov., sp. nov., obligate organohalide-respiring anaerobic bacteria, relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidetes* classis nov., within the phylum *Chloroflexi*. *International Journal of Systematic and Evolutionary Microbiology* **Apr 27** Epub ahead of print.

Mavromatis, K., Ivanova N. N., Chen I.-M. A., Szeto E., Markowitz V. M. and Kypides N. C. (2009). The DOE-JGI standard operating procedure for the annotations of microbial genomes. *Standards in Genomic Science* **1**: 63-67.

Maymó-Gatell, X., Chien Y.-t., Gosset J. M. and Zinder S. H. (1997). Isolation of a bacterium that reductively dechlorinates tetrachlorethene to ethene. *Science* **276**: 1568 - 1571.

McMurdie, P. J., Behrens S., Müller J. A., Göke J., Ritalahti K. M., Wagner R., et al. (2009). Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genetics* **5**: e1000714.

McMurdie, P.J., Hug, L., Edwards, E.A., Holmes, S., and Spormann, A.M. (2011) Site-specific mobilization of vinyl chloride respiration islands by a mechanism common in *Dehalococcoides*. *BMC Genomics* **12**: 287.

Men, Y., Feil H., VerBerkmoes N. C., Shah M. B., Johnson D. R., Lee P. K. H., et al. (2012). Sustainable syntrophic growth of *Dehalococcoides ethenogenes* strain 195 with *Desulfovibrio vulgaris* Hildenborough and *Methanobacterium congolense*: global transcriptomic and proteomic analyses. *The ISME Journal* **6**: 410-421.

Ni, S., Fredrickson J. K. and Xun L. (1995). Purification and characterization of a novel 3-chlorobenzoate-reductive dehalogenase from the cytoplasmic membrane of *Desulfomonile tiedjei* DCB-1. *Journal of Bacteriology* **177**: 5135-5139.

O'Toole, G. A. and Escalante-Semerena J. C. (1993). *cobU*-dependent assimilation of nonadenosylated cobinamide in *cobA* mutants of *Salmonella typhimurium*. *Journal of Bacteriology* **175**: 6328-6336.

Panek, H. and O'Brian M. R. (2002). A whole genome view of prokaryotic haem biosynthesis. *Microbiology* **148**: 2273-2282.

Pas, B. A. v. d., Jansen S., Dijkema C., Schraa G., Vos W. M. d. and Stams A. J. M. (2001). Energy yield of respiration on chloroaromatic compounds in *Desulfitobacterium dehalogenans*. *Applied and Environmental Microbiology* **67**: 3958-3963.

Raux, E., Schubert H. L., Roper J. M., Wilson K. S. and Warren M. J. (1999). Vitamin B_{12}: Insights into biosynthesis's mount improbable. *Bioinorganic Chemistry* **27**: 110-118.

Raux, E., Thermes C., Heathcote P., Rambach A. and Warren M. J. (1997). A role of *Salmonella typhimurium* cbiK in cobalamin (Vitamin B_{12}) and siroheme biosynthesis. *Journal of Bacteriology* **179**: 3202-3212.

Roessner, C. A. and Scott A. I. (2006). Fine-tuning our knowledge of the anaerobic route to cobalamin (Vitamin B_{12}). *Journal of Bacteriology* **188**: 7331-7334.

Roth, J. R., Lawrence J. G. and Bobik T. A. (1996). Cobalamin (coenzyme B_{12}): Synthesis and biological significance. *Annual Reviews of Microbiology* **50**: 137-181.

Sanford, R. A., Cole J. R. and Tiedje J. M. (2002). Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68**: 893-900.

Scholz-Muramatsu, H., Neumann A., Messmer M., Moore E. and Diekert G. (1995). Isolation and characterization of *Dehalospirillum multivorans* gen. nov. sp. nov., a tetrachloroethene-utilizing, strictly anaerobic bacterium. *Archives of Microbiology* **163**: 48-56.

Scott, A. I. and Roessner C. A. (2002). Biosynthesis of cobalamin (vitamin B_{12}). *Biochemical Society Transactions* **30**: 613-620.

Seshadri, R., Adrian L., Fouts D. E., Eisen J. A., Phillippy A. M., Methe B. A., et al. (2005). Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105-108.

Shelton, D. R. and Tiedje J. M. (1984). Isolation and partial characterization of bacteria in an anaerobic consortium that mineralizes 3-chlorobenzoic acid. *Applied and Environmental Microbiology* **48**: 840-848.

Smidt, H. and Vos W. M. d. (2004). Anaerobic microbial dehalogenation. *Annual Review of Microbiology* **58**: 43-73.

Sung, Y., Fletcher K. E., Ritalahti K. M., Apkarian R. P., Ramos-Hernández N., Sanford R. A., Mesba N. M. and Löffler F. E. (2006). *Geobacter lovleyi* sp. nov. strain SZ, a novel metal-reducing and tetrachloroethene-dechlorinating bacterium. *Applied and Environmental Microbiology* **69**: 2775-2782.

Teramoto, H., Suda M., Inui M. and Yukawa H. (2010). Regulation of the expression of genes involved in NAD de novo biosynthesis in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology* **76**: 5488-5495.

Thomas, M. G. and Escalante-Semerena J. C. (2000). Identification of an alternative nucleoside triphosphate:5'-deoxyadenosylcobinamide phosphate nucleotidyltransferase in *Methanobacterium thermoautotrophicum* DH. *Journal of Bacteriology* **182**: 4227-4233.

Thomas, S. H., Sanford R. A., Amos B. K., Leigh M. B., Cardenas E. and Löffler F. E. (2010). Unique ecophysiology among U(VI)-reducing bacteria as revealed by evaluation of oxygen metabolism in *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology* **76**: 176-183.

Thomas, S. H., Wagner R. D., Arakaki A. K., Skolnick J., Kirby J. R., Shimkets L. J., Sanford R. A. and Löffler F. E. (2008). The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* **3**: e2103.

Utkin, I., Woese C. and Wiegel J. (1994). Isolation and characterization of *Desulfitobacterium dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *International Journal of Systematic Bacteriology*. **44**: 612-619.

Villemur, R., Lanthier M., Beaudet R. and Lepine F. (2006). The *Desulfitobacterium* genus. *FEMS Microbiology Reviews* **30**: 706-733.

Wagner, D. D., Hug L. A., Hatt J. K., Spitzmiller M. R., Padilla-Crespo E., Ritalahti K. M., Edwards E. A., Konstantinidis K. T. and Löffler F. E. (2012). Genomic determinants of organohalide-respiration in *Geobacter lovleyi*, an unusual member of the *Geobacteraceae*. *BMC Genomics* **13**: 200.

Warren, M. J., Raux E., Schubert H. L. and Escalante-Semerena J. C. (2002). The biosynthesis of adenosylcobalamin (vitamin B_{12}). *Natural Product Reports* **19**: 390-412.

Woodson, J. D. and Escalante-Semerena J. C. (2004). CbiZ, an amidohydrolase enzyme required for salvaging the coenzyme B_{12} precursor cobinamide in *Archaea*. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 3591-3596.

Woodson, J. D., Zayas C. L. and Escalante-Semerena J. C. (2003). A new pathway for salvaging the coenzyme B_{12} precursor cobinamide in archaea requires cobinamide-phosphate synthase (CbiB) enzyme activity. *Journal of Bacteriology* **185**: 7193-7201.

Wu, Q., Sanford R. A. and Löffler F. E. (2006). Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology* **72**: 3608-3614.

Yan, J., Rash B. A., Rainey F. A. and Moe W. M. (2009). Isolation of novel bacteria within the *Chloroflexi* capable of reductive dechlorination of 1,2,3-trichloropropane. *Environmental Microbiology* **11**: 833-843.

Yan, J., Ritalahti K. M., Wagner D. D. and Löffler F. E. (2012). Unexpected Specificity of Interspecies Cobamide Transfer from *Geobacter* spp. to Organohalide-Respiring *Dehalococcoides mccartyi* strains. *Applied and Environmental Microbiology* **78**: 6630-6636.

Zhang, Y., Rodionov D. A., Gelfand M. S. and Gladyshev V. N. (2009). Comparative genomic analyses of nickel, cobalt, and vitamin B_{12} utilization. *BMC Genomics* **10**: 78.

CHAPTER 6

COMPUTATIONAL DESCRIPTION OF NITROUS OXIDE REDUCTASES IN ORGANOHALIDE-RESPIRING *δ-PROTEOBACTERIA.*

6.1 Abstract.

Nitrous oxide reductases (NosZ) of *Bacteria* and *Archaea* provide energy for anaerobic growth via the reduction of N_2O to inert dinitrogen (N_2), suggesting a natural N_2O sink is present throughout subsurface environments. Predicted *nosZ* genes have been found on the genomes of the chlorophenol-respiring *Anaeromyxobacter dehalogenans* strains 2CP-C, 2CP-1, and K, plus the chlorobenzoate-respiring *Desulfomonile tiedjei* strain DCB-1. Multiple sequence alignments (MSAs) and hidden Markov models (HMMs) indicate the *Anaeromyxobacter* and *Desulfomonile* NosZ likely represent functional nitrous oxide reductases. Phylogenetic reconstruction of 136 characterized and predicted NosZ show that *Anaeromyxobacter* and *Desulfomonile* NosZ comprise a sub-family of nitrous oxide reductases, "atypical NosZ", distinct from the extensively-characterized typical NosZ of Gram negative denitrifiers. Moreover, the atypical NosZ represented on genomes of organohalide-respiring *Bacteria* (OHRB), ammonifiers (e.g., *Wolinella succinogenes*), and gram positive denitrifiers (*Geobacillus thermodenitrificans*) is distributed across 13 Bacterial phyla and two Archaeal phyla. NosZ-encoding gene clusters (*nos*) from *Anaeromyxobacter* and *Desulfomonile* genomes exhibited a unique organization of genes predicted to encode electron-transport proteins and NosZ maturation proteins. Such insights into the diversity of *nosZ* genes and their novel features in *Anaeromyxobacter* and *Desulfomonile* may enable more accurate modeling of N_2O flux or environmental sequencing studies of N_2O respiring populations.

6.2 Introduction.

Prokaryotes play a major role in transforming nitrogen between bioavailable forms, such as nitrate (NO_3^-), and gaseous forms (N_2). Microbial processes may also prove

critical to the modeling or mitigation of human disturbances to the global nitrogen cycle. Denitrifiers couple anaerobic growth to the stepwise reduction of nitrate to nitrous oxide or nitrogen through the respiratory enzymes: Nap/Nar ($NO_3^- \rightarrow NO_2^-$), Nir ($NO_2^- \rightarrow NO$), Nor ($NO \rightarrow N_2O$), and NosZ ($N_2O \rightarrow N_2$) (Zumft, 1997). By comparison, ammonifiers couple anaerobic growth to the reduction of nitrate to ammonia ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$) (Simon, 2002) while nitrifiers catalyze a reverse processes ($NH_4^+ \rightarrow NO_2^-$) under aerobic conditions (Wrage et al., 2001). Each of these three groups of NOx-respiring microbes are implicated in emissions of nitrous oxide, N_2O, but denitrifiers are believed to play the dominant role in N_2O flux (Conrad, 1996; Zumft, 1997). Increasing atmospheric N_2O concentrations over the past two centuries can be attributed to industrial nitrogen fixation via the Haber-Bosch process; In turn, massive influxes of fixed nitrogen through inorganic fertilizers has stimulated the activity of denitrifiers (Machida et al., 1995; Conrad, 1996; Zumft, 1997; Galloway, 2004). Nitrous oxide exerts greenhouse gas effects over 300 times more potent than CO_2 (Finlayson-Pitts et al., 2000; Lashof et al., 1990; Wang et al., 1976; Forster et al., 2007) and depletes stratospheric ozone (Cicerone et al., 1987; Crutzen et al., 1970; Ravishankara et al., 2009). Hence, the sustainability, or even habitability, of Earth may depend upon a more complete understanding of microbes involved in N_2O flux.

While it is unclear whether microbial processes serve to increase or decrease net atmospheric N_2O, the nitrous oxide reductase enzyme, NosZ, transforms the reactive N_2O to inert N_2 and is found only among *Bacteria* and some *Archaea* (Zumft and Körner, 2007). NosZ proteins characterized by biochemical studies or crystal

structures exhibit a two-domain structure, where the N-terminal domain comprises a seven-bladed beta propeller coordinating a copper-sulfur catalytic center of novel 3D structure (Haltia et al., 2003; Paraskevopoulos et al., 2006; Pomowski et al., 2011). Moreover, the NosZ beta propeller domain contains eight histidine residues conserved across all experimentally-characterized NosZ from denitrifiers such as *Pseudomonas stutzeri*, *Bradyrhiobium japonicum*, *Achromobacter cycloclastes*, as well as the ammonifier, *Wolinella succinogenes* (Riester et al., 1989; Velasco et al., 2004; Paraskevopoulos et al., 2006; Simon et al., 2004). Given that these sequence features are found in no other known protein, *nosZ* gene homologs, presumably encoding functional NosZ proteins, can be identified via bioinformatics or molecular assays (e.g., PCR) with confidence.

The increasing availability of genomes from microbes of subsurface environments reveals a greater diversity of organisms capable of N_2O to N_2 transformation than previously characterized among denitrifiers. Members of the *Anaeromyxobacter* genus are able to respire upon a remarkable range of terminal electron acceptors and likely play a crucial role in nitrogen cycling. The organohalide-respiring *A. dehalogenans* strains 2CP-C, 2CP-1 and K, isolated from rain forest soil, suburban compost, and freshwater sediments (Sanford et al., 2002), have been characterized for their ability to respire upon hexavalent uranium (Wu et al., 2006), 2-chlorophenol (Sanford et al., 2002), nitrate/nitrite (Sanford et al., 2002), and sub-atmospheric concentrations of oxygen (Thomas et al., 2010). Sequencing and analysis of the *A. dehalogenans* 2CP-C genome revealed a gene predicted to encode NosZ (Thomas et al., 2008), suggesting the strain is able to utilize N_2O as a terminal electron acceptor. As no genetic system exists for members of the *Anaeromyxobacter* genus, the current study investigates the evolutionary history and functionality of *nosZ* on

four available *Anaeromyxobacter* genomes and on the genome of the related chlorobenzoate-respirer, *Desulfomonile tiedjei.*

6.3 Results and Discussion.

6.3.1 Evidence that *δ-Proteobacteria* NosZ function in N_2O respiration.

Predicted NosZ from *Anaeromyxobacter* and *Desulfomonile* shared functionally-critical amino acid sequence features with genetically- and structurally-characterized NosZ proteins, thus providing evidence for NosZ functionality in *δ-Proteobacteria*. A multiple sequence alignment (MSA) revealed the *Anaeromyxobacter* NosZ shared all seven of the conserved histidine residues (Fig. 1 A, B, and C, red sites) in the NosZ beta propeller domain shown to coordinate the Cu_Z-center in NosZ crystal structures (Haltia et al., 2003; Paraskevopoulos et al., 2006; Pomowski et al., 2011). By contrast, *δ-Proteobacteria* NosZ shared only two histidine residues with the copper-binding nitrite reductase, NirK, from *Achromobacter cycloclastes*. *δ-Proteobacteria* NosZ possessed four out of the six calcium-binding residues shown to mediate NosZ dimerization (Fig. 1, cyan sites, Haltia et al., 2003) while C-terminal regions of the *δ-Proteobacteria* NosZ shared five out of the six Cu_A-binding residues with characterized NosZ (Fig. 1 C, green sites). A hidden Markov model (HMM) constructed from aligned beta propeller domains of structurally-characterized NosZ (Fig. 1) retrieved NosZ only from loci on genomes of *Anaeromyxobacter* strains 2CP-C, 2CP-1, K, and Fw109-5 (Adeh_2402, A2cp1_1556, AnaeK_1461, and Anae109_0244) and *Desulfomonile tiedjei* (Desti_0657) identified in previous Blast analyses. *Anaeromyxobacter* pure cultures have exhibited growth with N_2O as the sole terminal electron acceptor, providing supporting evidence that the predicted NosZ are

functional (Sanford et al., in press). Hence, MSA and HMM analyses provide evidence that *Anaeromyxobacter* and *Desulfomonile* genomes each harbor a single *nosZ* gene likely to encode a functional N_2O reductase.

Despite sequence features shared with characterized NosZ from denitrifiers of the α-, β-, and γ-*Proteobacteria*, δ-*Proteobacteria* NosZ clustered with a separate NosZ sub-family, represented by the NosZ from the ammonifier, *Wolinella succinogenes*. An MSA of 136 predicted and characterized NosZ from completed and draft genomes revealed that 77 NosZ shared a key set of sequence features with the seven experimentally-characterized NosZ at the top of the alignment in Fig. 1 and were termed "typical NosZ". All 77 typical NosZ shared pairwise amino acid identities > 40% (avg. 58%) and represented genomes of the α-, β-, and γ-*Proteobacteria*. Moreover, amino acid sequence features conserved in typical NosZ exhibited less conservation in the NosZ of *Anaeromyxobacter* and *Desulfomonile*. In the beta propeller domain, three of the Cu_Z-binding histidine residues comprised 100% conserved motifs **D**x**<u>HH</u>**x**H** and **EP<u>H</u>D** across typical NosZ, while only the central histidine sites (underlined) were conserved in *Anaeromyxobacter* and *Desulfomonile* NosZ (Fig. 1 A and C). All typical NosZ from complete or draft genomes possessed an N-terminal twin-arginine-translocation (Tat-type) signal peptide with the **RR**x(**F**|**L**) motif resembling the signal implicated in NosZ export to the periplasm in *Pseudomonas stutzeri* (Berks, 2005; Heikkilä et al., 2001). By contrast, *Anaeromyxobacter* and *Desulfomonile* NosZ lacked recognizable Tat motifs, instead possessing a Sec-type N-terminal signal peptide (den Blaawen and Driessen, 1996). The *Anaeromyxobacter* and *Desulfomonile* NosZ protein sequences did share strong similarity with genetically-characterized NosZ from the ammonifying ε-

Proteobacterium, *Wolinella succinogenes* (Payne et al., 1982; Simon et al., 2004) and the Gram-positive denitrifier, *Geobacillus thermodenitrificans* (Liu et al., 2008). Like the NosZ of *δ-Proteobacteria*, *Wolinella* and *Geobacillus* NosZ exhibit a Sec-type N-terminal signal peptide. Unlike the *δ-Protebacteria* NosZ, the *Wolinella succinogenes* NosZ protein has a *c*-type cytochrome domain at its C-terminus (Simon et al., 2004). Nonetheless, *Anaeromyxobacter* NosZ share 39% and 53% amino acid identity with NosZ from *Wolinella succinogenes* and *Geobacillus thermodenitrificans*, respectively. By contrast, *Anaeromyxobacter* NosZ shares only 33% identity with the typical NosZ from the model soil denitrifier, *Bradyrhizobium japonicum* (Velasco et al., 2004). Hence, comparative sequence analyses allowed *Anaeromyxobacter* and *Desulfomonile* NosZ to be grouped with *Wolinella* and *Geobacillus* NosZ, together termed "atypical NosZ" in contrast with the typical NosZ of model Gram negative denitrifiers. In fact, pairwise sequence alignments between 77 typical NosZ and 55 atypical NosZ revealed that the two sub-families consistently share less than 40% amino acid identity (avg. 32%). Hence, atypical NosZ including *δ-Proteobacteria* NosZ, exhibited wider sequence diversity than typical NosZ, possibly a consequence of varying taxonomic affiliation or ecophysiological conditions.

6.3.2 Phylogenetic and ecological distribution of NosZ

Phylogenetic analyses of 136 NosZ from 133 completed and draft genomes provided evidence that atypical NosZ proteins represent a greater diversity of taxonomic groups, habitats, and NOx-respiratory pathways than has been previously characterized for typical NosZ. In the 136-sequence phylogeny, the atypical NosZ cluster containing the *Anaeromyxobacter* and *Desulfomonile* was comprised of 55 sequences distributed

across 54 genomes representing 13 phyla of *Bacteria* and two phyla of *Archaea*. A majority of atypical NosZ, 52 out of the 55 total, exhibited a Sec-type N-terminal signal peptide (den Blaawen and Driessen, 1996) rather than the N-terminal Tat motif. Only atypical NosZ of deep-branching taxonomic groups, such as the Archaeon, *Ferroglobus placidus*, and the *Chloroflexi*, *Sphaerobacter thermophilus* and *Thermomicrobium roseum*, were found to have the Tat-type signal (**RR**x(**F**|**L**)). C-terminal regions homologous to the *c*-type cytochrome domain of the atypical NosZ protein from *Wolinella succinogenes* were found only in NosZ of *ε-Proteobacteria* and *Deferribacteres* (Fig. 2). Thus, N- and C-terminal variability among atypical NosZ indicates an ancestral evolutionary relationship with multiple microbial phyla and/or extensive sequence divergence across habitats in contrast with the taxonomically-restricted typical NosZ.

Investigation of genomic affiliations for atypical NosZ revealed associations with denitrifiers, ammonifiers, and various extremophiles, as well as with metabolically-versatile OHRB such as *Anaeromyxobacter dehalogenans*. NosZ from the chlorophenol-respiring *Anaeromyxobacter dehalogenans* strains clustered with the NosZ from the non-organohalide-respiring *Anaeromyxobacter* sp. strain Fw109-5 and the chlorobenzoate-respiring *Desulfomonile tiedjei*, comprising a single clade of *δ-Proteobacteria* NosZ (Fig. 1, top). Atypical NosZ were also found on four genomes of the organohalide-respiring *Desulfitobacterium* spp. Moreover, strains of *Anaeromyxobacter* and *Desulfitobacterium* were among 15 genomes with atypical NosZ which lacked NirK or NirS, but possessed ammonia-forming nitrite reductases, NrfA. Up to 27% of genomes with atypical NosZ thus harbored genes enabling ammonification, ($NO_2^- \rightarrow NH_4^+$), including the genome of the model ammonifier, *Wolinella succinogenes* (Simon et al., 2004). Twenty-four genomes harboring atypical

NosZ also possessed *nirK* or *nirS*, indicating the ability to carry out full denitrification. Additional genomes harboring atypical *nosZ* with denitrification-related *nirK* or *nirS* represented extremophilic phyla such as the deep-branching *Bacteria*, *Gemmatimonas aurantiaca*, *Hydrogenobacter thermophilus*, *Sphaerobacter thermophilus*, and the Archaeon, *Pyrobaculum calidifontis* (Fig. 1). Two strains harboring atypical *nosZ*, *Opitutus terrae* and *Marivirga tractuosa*, plus *Shewanella loihica*, with typical *nosZ*, each harbored *nirK* and *nrfA* genes, suggesting denitrification and ammonification may be compatible pathways (Sanford et al., in press). Finally, 16 genomes with atypical NosZ lacked all types of nitrite reductases, including the OHRB, *Desulfomonile tiedjei*, the halophilic *Salinibacter ruber*, and the thermophiles, *Thermomicrobium roseum*, and *Ferroglobus placidus* (Fig. 1). In summary, the distribution of atypical NosZ on genomes of dentrifiers, ammonifiers, and organisms unable to use nitrite indicates N_2O reduction is not strictly linked with denitrification. The presence of NosZ across deeply-branched lineages of extremophiles also suggests N_2O reduction is a widespread and physiologically versatile process; or alternatively, not all NosZ may actually function in N_2O reduction.

By contrast, the 77 typical NosZ sequences were found only on genomes of α-, β-, and *γ-Proteobacteria*, 75 strains represented in all (Fig. 2). Moreover, only ten of the 75 genomes harboring typical *nosZ* genes lacked genes encoding the key denitrifier enzymes, NO-forming nitrite reductases, NirS or NirK (Fig. 1). In this manner, the typical NosZ were predominantly associated *Bacteria* confirmed or predicted to catalyze full denitrification.

6.3.3 NosZ gene cluster organization.

Gene clusters harboring *nosZ* are part of the *nos* cluster, which includes accessory genes required for maturation and respiratory function of the NosZ protein. The *Anaeromyxobacter nos* clusters comprise 10 or 11 genes while the *Desulfomonile tiedjei nos* cluster comprises 12 genes (Fig. 3). *nos* clusters from both *δ-Proteobacteria* genera share the four NosZ maturation genes *nosD*, *nosF*, *nosY*, and *nosL* with *nos* clusters harboring typical *nosZ* (Fig. 3, blue-shaded genes). *δ-Proteobacteria nos* clusters, as well as all other *nos* clusters with atypical *nosZ*, lack *nosR* and *nosX*, believed to function in electron transfer to NosZ during maturation or respiration (Cuypers et al., 1992; Philippot, 2002; Honisch et al., 2003). The absence of *nosR* and *nosX* from *nos* clusters harboring experimentally-characterized *nosZ* in *Wolinella succinogenes* and *Geobacillus thermodenitrificans* suggests these genes are not essential to the functionality of atypical NosZ (Fig. 3). Hence, despite differences in gene organization, the conservation of the core *nos* genes, *nosZ*, *nosD*, *nosF*, *nosY*, and *nosL*, with characterized *nos* clusters provides further evidence of *δ-Proteobacteria nosZ* functionality.

The *Anaeromyxobacter* and *Desulfomonile nos* clusters all shared five genes in common which encoded functions distinct from those found in typical *nos* clusters. In these *δ-Proteobacteria nos* clusters, four genes upstream of *nosZ* encode proteins with 4Fe-4S iron sulfur motifs (**C**xx**C**xx**C**xxx**CP),** a 2Fe-2S Rieske iron-sulfur motif (**C**x**H**…**CPCH**), two *b*-type cytochrome domains (pfam0003(2|3)), and five *c*-type cytochrome (**C**xx**CH**) motifs, respectively, suggesting they serve electron transport functions (Fig. 3). Electron-transport-related functions were found upstream of *nosZ* in the *Desulfitobacterium nos* clusters, yet exhibited variable organization in other *nos*

clusters harboring atypical nosZ (Fig. 3). For example, the *nos* cluster of *Wolinella succinogenes*, encoding a NosZ with a C-terminal *c*-type cytochrome, encodes iron-sulfur proteins and additional *c*-type cytochromes downstream of *nosZ*. A gene encoding a transmembrane protein of unknown function (Fig. 3, gene 'TM'), typically situated downstream of *nosZ*, was found in all *nos* clusters with typical *nosZ*. In summary, the novel gene organization shared by *Anaeromyxobacter* and *Desulfomonile nos* clusters suggests regulatory coupling or protein-protein interactions critical to NosZ function in these organisms.

6.4 Conclusions and implications.

The current study has shown that metabolically-versatile OHRB may be crucial to the geochemical cycling of greenhouse gases, such as N_2O, in addition to contaminants of soil and groundwater such as organohalides. PCR data indicate that *Anaeromyxobacter* strains and atypical *nosZ* genes similar to those found in *Anaeromyxobacter* and *Desulfomonile* are widespread in agricultural soils (Sanford et al., in press). The bioinformatics methodology of this study elucidated features of atypical *nosZ* which may facilitate the development of molecular tools, such as PCR or microarrays, yielding a more complete picture of N_2O flux in soils. It is generally believed that denitrifiers play the dominant role in N_2O flux to the atmosphere (Conrad, 1996; Cuhel et al., 2010; Wallenstein et al., 2006; McSwinney and Robertson, 2005); Yet the current phylogenetic analyses indicate that ammonifiers with the ability to respire N_2O, such as *Anaeromyxobacter*, should be considered for incorporation into N_2O emission models. Furthermore, *Anaeromyxobacter* and *Desulfomonile* have been extensively characterized for their potential applications to contaminant mitigation in the subsurface and in wastewater sludges, respectively

(Sanford et al., 2002; Wu et al., 2006; Thomas et al., 2009; Mohn and Kennedy, 1992; Ahring et al., 1992). Hence, the viability of *Anaeromyxobacter* and *Desulfomonile* as models for N_2O conversion in agricultural or waste treatment processes should likewise be investigated.

6.5 Methods.

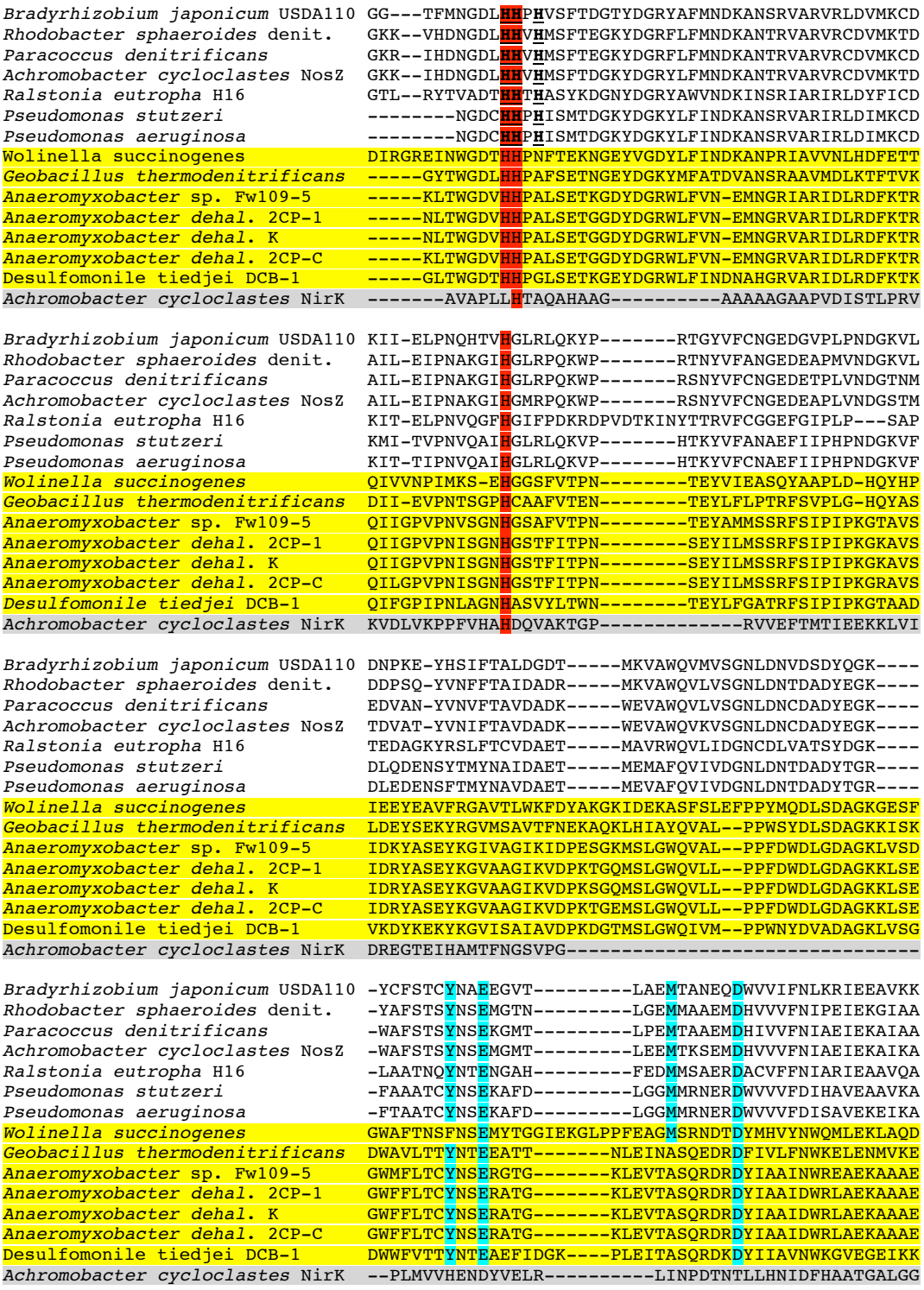
A set of 136 NosZ amino acid sequences, representing 133 complete genome sequences, were downloaded from GenBank. Sequences of functionally characterized NosZ proteins and their accessory genes from *Bradyrhizobium japonicum* USDA110 (NP_766955) and *Wolinella succinogenes* (NC_005090), plus sequences corresponding to NosZ with resolved crystal structures, 1FWX of *Paracoccus denitrificans* and 2IWF of *Achromobacter cycloclastes*, comprised a hidden Markov model in HMMer-3.0 (Eddy, 1998) to query GenBank or RefSeq genome sequences. Accessory genes of NosZ from the genomes of *B. japonicum* strain USDA110 and *W. succinogenes* were used as PSI-Blast queries (Altschul et al., 1997) to analyze genes adjacent to inferred *nosZ* on *Anaeromyxobacter* genomes. Sequences of functionally characterized NirK and NirS, plus sequences corresponding to NrfA with resolved crystal structures, comprised hidden Markov models to query NosZ-encoding genomes to infer either a denitrification or DNRA gene repertoire. Retrieval of characterized N-metabolism sequences from experimentally characterized denitrifiers or ammonifiers validated the HMM approach (e.g., NosZ from *Geobacillus thermodenitrificans* with e-value = 9e-126). Predicted NosZ, NirK, NirS, and NrfA were distinguished from possible false positives by presence of conserved amino acid residues and protein domains consistent with those represented in the NosZ, NirK/NirS, and NrfA HMMs. Freely available online tools for parsing conserved domains (PFAM - pfam.janelia.org), transmembrane helices (TMHMM2 - smart.embl-heidelberg.de),

and signal peptides (TatP, www.cbs.dtu.dk/services/TatP; SignalP, www.cbs.dtu.dk/services/SignalP) were used to determine similarities between NosZ and other *nos* cluster genes. *nos* gene clusters visualized with Artemis (www.sanger.ac.uk/resources/software) guided the graphic display of *nos*-cluster gene diagrams (www.gimp.org). A multiple sequence alignment was generated from 136 NosZ sequences from closed NCBI genomes using Muscle (Edgar, 2004). From this alignment, neighbor-joining consensus trees were constructed in Phylip (Felsenstein, 1989) based upon 200 bootstrapped replicates of NosZ amino acid sequence alignments.

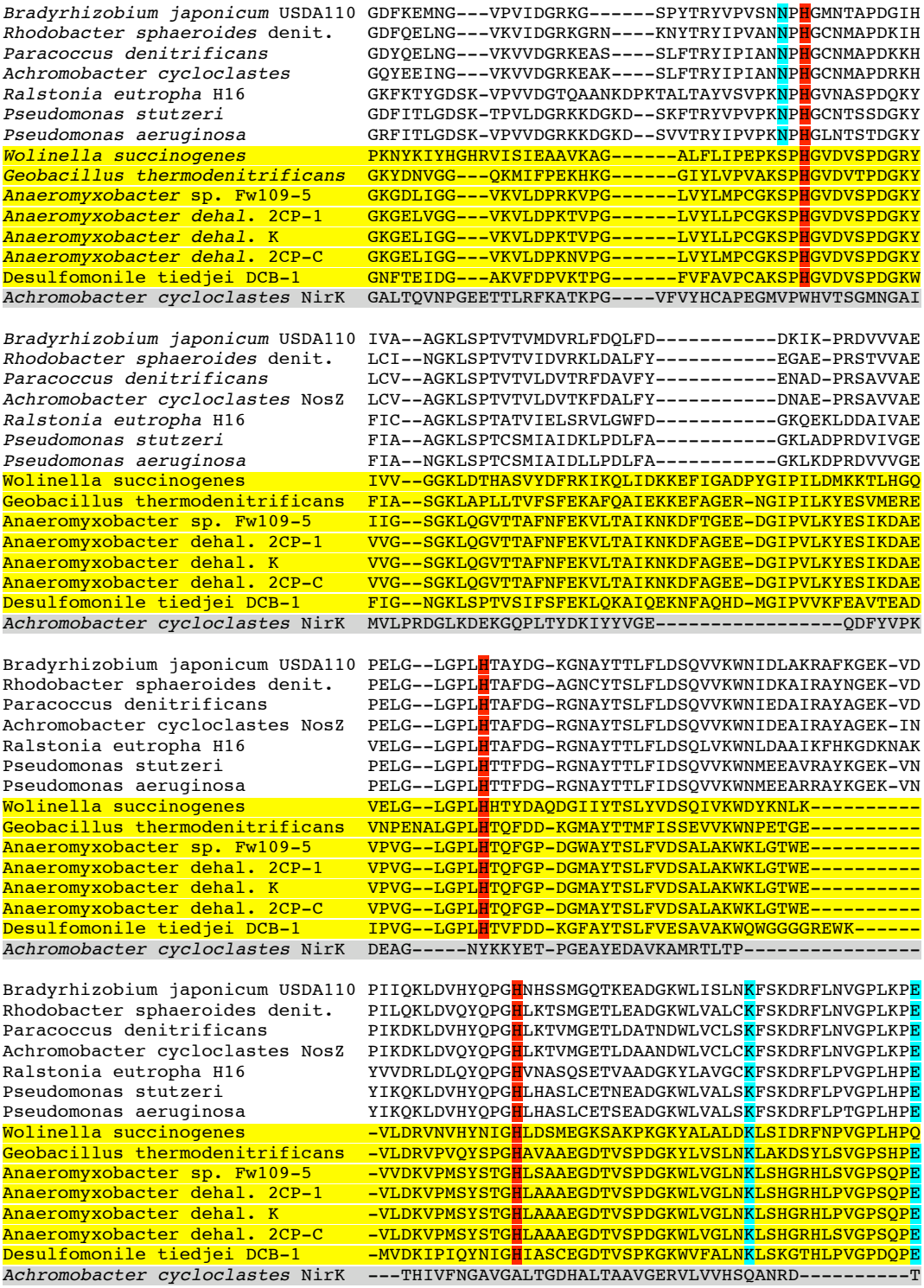
6.6 Chapter 6 Acknowledgements.

This research was supported by the U.S. DOE Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program, Award No. DESC0006662.

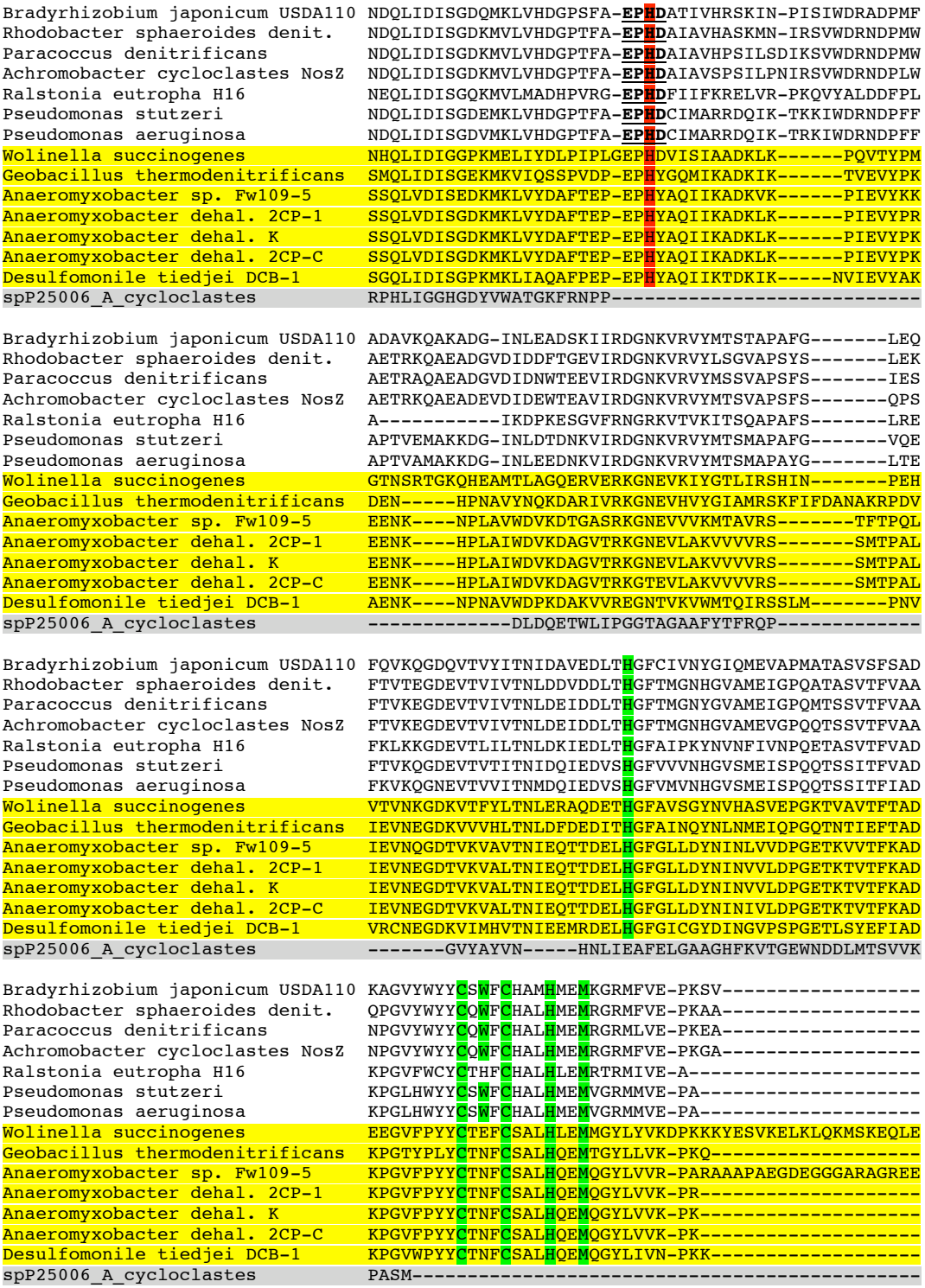
6.7 Chapter 6, Figures.



6.7.1 Figure 1 A. Atypical NosZ (yellow) and NirK (gray) with Cu_Z histidine ligands to the (red) and residues mediating NosZ dimerization (cyan). Underlined residues at first histidine ligand serve to distinguish typical NosZ from atypical NosZ.

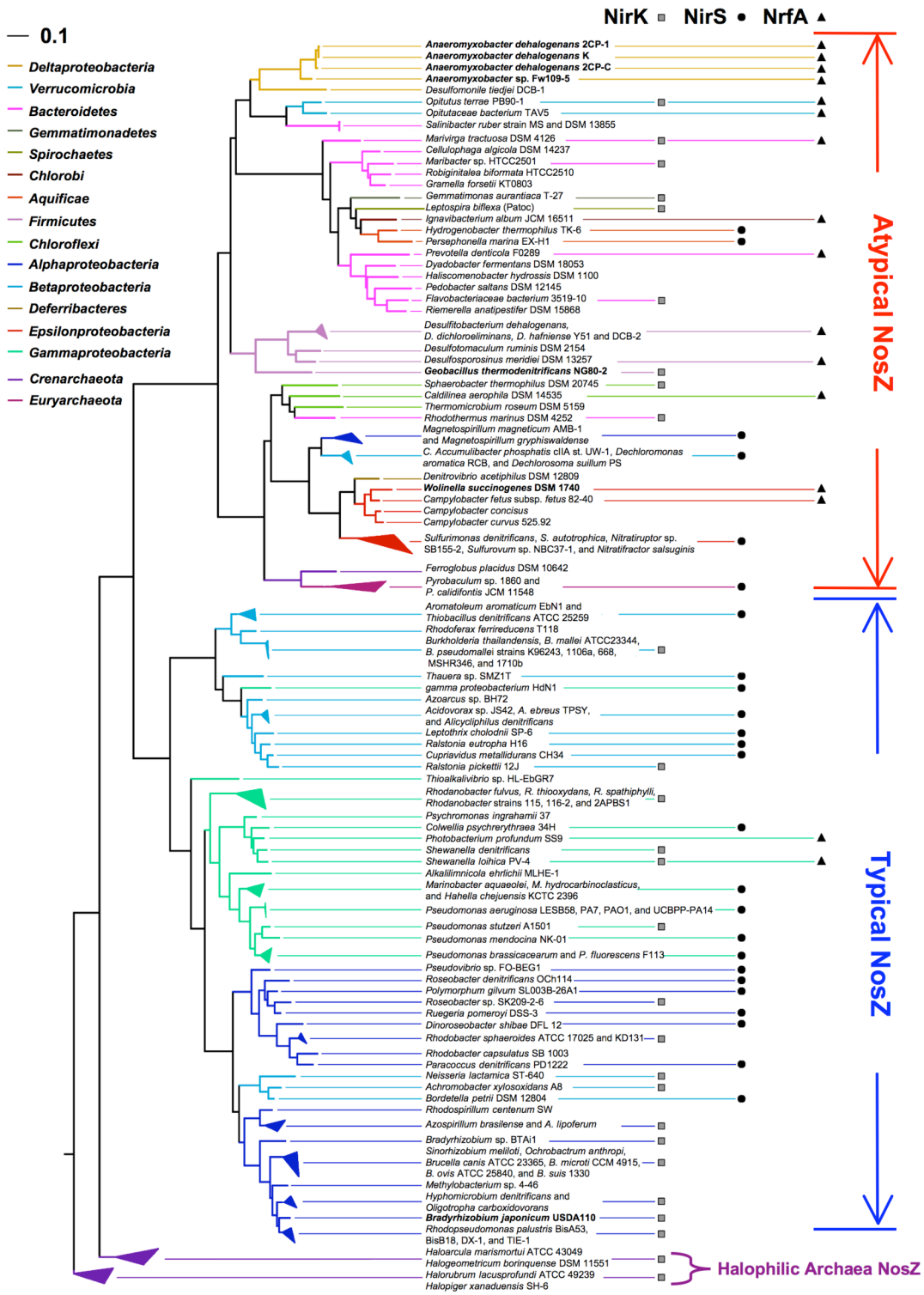


6.7.1 Figure 1 B. Atypical NosZ (yellow) and NirK (gray) with first two histidine ligands to the Cu_Z center (red) and three conserved residues mediating NosZ dimerization (cyan).

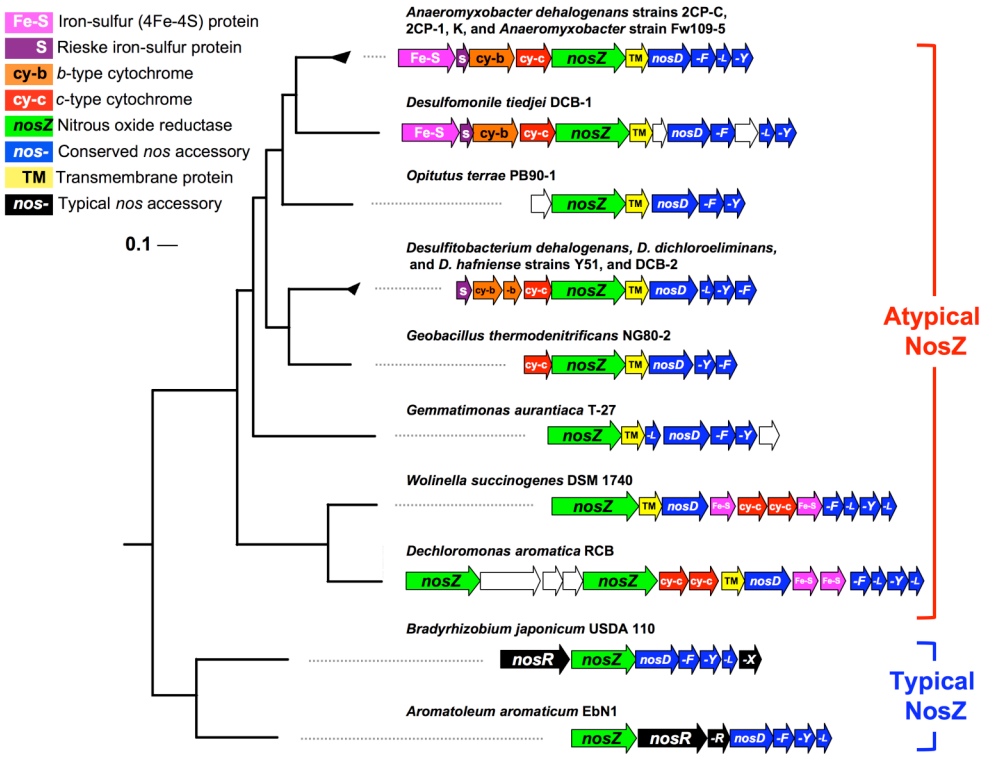


6.7.1 Figure 1 C. C-terminal cupredoxin domain of NosZ where residues binding the Cu_A-center are highlighted in green. Underlined residues adjacent to the seventh histidine ligand (red) serve to distinguish typical NosZ from atypical NosZ.

6.7.1 Figure 1. A, B, and C. *Anaeromyxobacter* and *Desulfomonile* NosZ aligned with experimentally characterized NosZ and NirK protein sequences. Included in the alignment are the genetically and/or biochemically characterized Z-type (typical) NosZ, sp|Q89XJ6, sp|AAD43473, sp|Q01710, and sp|Q59105 from *Bradyrhizobium japonicum* USDA110, *Rhodobacter sphaeroides* subsp. *denitrificans*, *Pseudomonas aeruginosa*, and *Ralstonia eutropha* H16, respectively. NosZ with solved crystal structures, 1FWX, 2IWF, and 3SBP, are from *Paracoccus denitrificans*, *Achromobacter cycloclastes*, and *Pseudomonas stutzeri*, respectively. Highlighted in yellow are the atypical NosZ from *Desulfomonile tiedjei* strain DCB-1 (Desti_0657), *Anaeromyxobacter* sp. strain Fw109-5 (Anae109_0244), *Anaeromyxobacter dehalogenans* strain 2CP-1 (A2cp1_1556), strain K (AnaeK_1461), and strain 2CP-C (Adeh_2402), as well as the characterized atypical NosZ from *Wolinella succinogenes* (CAG26676) and *Geobacillus thermodenitrificans* (GTNG_1734). To highlight the sequence uniqueness of NosZ, a region of the characterized NirK (sp|P25006) from *Achromobacter cycloclastes*, which exhibited matches to NosZ sequences in Blast, is included in the alignment (highlighted in gray).


NirK
NirS
NrfA
0.1
Deltaproteobacteria
Verrucomicrobia
Bacteroidetes
Gemmatimonadetes
Spirochaetes
Chlorobi
Aquificae
Firmicutes
Chloroflexi
Alphaproteobacteria
Betaproteobacteria
Deferribacteres
Epsilonproteobacteria
Gammaproteobacteria
Crenarchaeota
Euryarchaeota
Anaeromyxobacter dehalogenans 2CP-1
Anaeromyxobacter dehalogenans K
Anaeromyxobacter dehalogenans 2CP-C
Anaeromyxobacter sp. Fw109-5
Desulfomonile tiedjei DCB-1
Opitutus terrae PB90-1
Opitutaceae bacterium TAV5
Salinibacter ruber strain MS and DSM 13855
Marivirga tractuosa DSM 4126
Cellulophaga algicola DSM 14237
Maribacter sp. HTCC2501
Robiginitalea biformata HTCC2510
Gramella forsetii KT0803
Gemmatimonas aurantiaca T-27
Leptospira biflexa (Patoc)
Ignavibacterium album JCM 16511
Hydrogenobacter thermophilus TK-6
Persephonella marina EX-H1
Prevotella denticola F0289
Dyadobacter fermentans DSM 18053
Haliscomenobacter hydrossis DSM 1100
Pedobacter saltans DSM 12145
Flavobacteriaceae bacterium 3519-10
Riemerella anatipestifer DSM 15868
Desulfitobacterium dehalogenans,
D. dichloroeliminans, D. hafniense Y51 and DCB-2
Desulfotomaculum ruminis DSM 2154
Desulfosporosinus meridiei DSM 13257
Geobacillus thermodenitrificans NG80-2
Sphaerobacter thermophilus DSM 20745
Caldilinea aerophila DSM 14535
Thermomicrobium roseum DSM 5159
Rhodothermus marinus DSM 4252
Magnetospirillum magneticum AMB-1
and Magnetospirillum gryphiswaldense
C. Accumulibacter phosphatis clIA st. UW-1, Dechloromonas
aromatica RCB, and Dechlorosoma suillum PS
Denitrovibrio acetiphilus DSM 12809
Wolinella succinogenes DSM 1740
Campylobacter fetus subsp. fetus 82-40
Campylobacter concisus
Campylobacter curvus 525.92
Sulfurimonas denitrificans, S. autotrophica, Nitratiruptor sp.
SB155-2, Sulfurovum sp. NBC37-1, and Nitratifractor salsuginis
Ferroglobus placidus DSM 10642
Pyrobaculum sp. 1860 and
P. calidifontis JCM 11548
Aromatoleum aromaticum EbN1 and
Thiobacillus denitrificans ATCC 25259
Rhodoferax ferrireducens T118
Burkholderia thailandensis, B. mallei ATCC23344,
B. pseudomallei strains K96243, 1106a, 668,
MSHR346, and 1710b
Thauera sp. SMZ1T
gamma proteobacterium HdN1
Azoarcus sp. BH72
Acidovorax sp. JS42, A. ebreus TPSY,
and Alicycliphilus denitrificans
Leptothrix cholodnii SP-6
Ralstonia eutropha H16
Cupriavidus metallidurans CH34
Ralstonia pickettii 12J
Thioalkalivibrio sp. HL-EbGR7
Rhodanobacter fulvus, R. thiooxydans, R. spathiphylli,
Rhodanobacter strains 115, 116-2, and 2APBS1
Psychromonas ingrahamii 37
Colwellia psychrerythraea 34H
Photobacterium profundum SS9
Shewanella denitrificans
Shewanella loihica PV-4
Alkalilimnicola ehrlichii MLHE-1
Marinobacter aquaeolei, M. hydrocarbinoclasticus,
and Hahella chejuensis KCTC 2396
Pseudomonas aeruginosa LESB58, PA7, PAO1, and UCBPP-PA14
Pseudomonas stutzeri A1501
Pseudomonas mendocina NK-01
Pseudomonas brassicacearum and P. fluorescens F113
Pseudovibrio sp. FO-BEG1
Roseobacter denitrificans OCh114
Polymorphum gilvum SL003B-26A1
Roseobacter sp. SK209-2-6
Ruegeria pomeroyi DSS-3
Dinoroseobacter shibae DFL 12
Rhodobacter sphaeroides ATCC 17025 and KD131
Rhodobacter capsulatus SB 1003
Paracoccus denitrificans PD1222
Neisseria lactamica ST-640
Achromobacter xylosoxidans A8
Bordetella petrii DSM 12804
Rhodospirillum centenum SW
Azospirillum brasilense and A. lipoferum
Bradyrhizobium sp. BTAi1
Sinorhizobium meliloti, Ochrobactrum anthropi,
Brucella canis ATCC 23365, B. microti CCM 4915,
B. ovis ATCC 25840, and B. suis 1330
Methylobacterium sp. 4-46
Hyphomicrobium denitrificans and
Oligotropha carboxidovorans
Bradyrhizobium japonicum USDA110
Rhodopseudomonas palustris BisA53,
BisB18, DX-1, and TIE-1
Haloarcula marismortui ATCC 43049
Halogeometricum borinquense DSM 11551
Halorubrum lacusprofundi ATCC 49239
Halopiger xanaduensis SH-6
Atypical NosZ
Typical NosZ
Halophilic Archaea NosZ

6.7.2 Figure 2. Bootstrapped neighbor-joining phylogeny of 136 NosZ sequences representing 133 genomes. Branches with bootstrap support <50% are indicated by open circles. The typical NosZ clade is comprised of sequences with an N-terminal Tat motif and associated primarily with functionally characterized soil denitrifiers of the *Proteobacteria*. The atypical NosZ is distributed among *Bacteria* and *Archaea*, and has been characterized in *A. dehalogenans*, *W. succinogenes*, and *Geobacillus thermodenitrificans* (bold font). *nosZ* occurs as a single copy gene on all genomes, except for *Pseudomonas brassicacearum* , *Pseudomonas fluorescens*, *Dechloromonas aromatica*, and *Sulfurimonas denitrificans*, which both have two *nosZ* loci. Also indicated is the presence of nitrite respiratory genes *nirS*/*nirK* (denitrification) and *nrfA* (DNRA ammonification). Four NosZ within the halophilic *Euryarchaeota* formed a discrete cluster and did not group with either the atypical or the typical NosZ. The scale bar at the top left corresponds to the mean number of amino acid substitutions per site.



6.7.3 Figure 3. Phylogenetic comparison of *nos* clusters. Comparison of *nos* clusters carrying the atypical *nosZ* with the characterized *nos* cluster of the complete denitrifier *Bradyrhizobium japonicum* strain USDA110, which harbors a typical *nosZ*. *nos* clusters of *Anaeromyxobacter* sp. strain Fw109-5, three *A. dehalogenans* strains, *Opitutus terrae*, *Desulfomonile tiedjei*, *Geobacillus thermodenitrificans*, four representatives of the *Desulfitobacterium genus*, *Gemmatimonas aurantiaca*, *W. succinogenes*, and *Dechloromonas aromatica* harbor the atypical *nosZ* and encode predicted iron-sulfur-binding proteins (labeled '4Fe-4S' or 'Fe-S'), Rieske iron-sulfur proteins (S), *b*-type cytochromes (cy-b), or *c*-type cytochromes. The gene encoding a protein with up to four predicted transmembrane-spanning regions (TM) is found in all atypical, but none of the typical, *nos* clusters. Accessory genes (*nosD*, *nosF*, *nosL*, and *nosY*) are generally conserved across *nos* clusters with both typical and atypical

nosZ. Non-colored genes in the *Opitutus terrae*, *Gemmatimonas aurantiaca,* and *Dechloromonas aromatica* clusters have no orthologs in any other known *nos* cluster. *nosR* and *nosX* are exclusively associated with typical *nos* clusters. *W. succinogenes* produces a unique NosZ with an additional C-terminal *c*-type cytochrome domain. The scale bar indicates the mean number of amino acid substitutions per site. Fe-S, iron-sulfur (4Fe-4S) protein; S, Rieske iron-sulfur protein; cy-b, *b*-type cytochrome; cy-c, *c*-type cytochrome; nos-, NosZ and accessory proteins; TM, transmembrane protein.

6.8 References.

Ahring, B.K., Christiansen, N., Mathrani, I., Hendriksen, H.V., Macario, A.J.L., and Macario, E.C. (1992) Introduction of a *de novo* bioremediation ability, aryl reductive dechlorination, into anaerobic granular sludge by inoculation with *Desulfomonile tiedjei*. *Applied and Environmental Microbiology*. **58**: 3611-3682.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25:**3389–3402.

Berks, B.C., Palmer, T., and Sargent, F. (2005) Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Curremt Opinion in Microbiology* **8**:174-181.

Bouwman, A.F., Van der Hoek, K.M., and Olivier, J.G.J. (1995) Uncertainties in the global source distribution of nitrous oxide. *Journal of Geophysical Research* **100**:2785-2800.

den Blaauwen, T. and Driessen, A.J.M. (1996) Sec-dependent preprotein translocation in bacteria. *Archives of Microbiology* **165**:1-8.

Cicerone, R.J. (1987) Changes in stratospheric ozone. *Science* **237**:35-42.

Conrad, R. (1996) Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O, and NO). *Microbiology Reviews* **60**: 609-640.

Crutzen, P.J. (1970) Influence of nitrogen oxides on atmospheric ozone content. *Quarterly Journal of the Royal Meteorological Society* **96**:320-325.

Cuypers, H., Viebrock-Sambale, A., and Zumft, W. (1992) NosR, a membrane-bound regulatory component necessary for expression of nitrous oxide reductase in denitrifying *Pseudomonas stutzeri*. *Journal of Bacteriology* **174**:5332-5339.

Čuhel, J., Simek, M., Laughlin, R.J., Bru, D., Chèneby, D., Watson, C.J., and Philippot, L. (2010) Insights into the effect of soil pH on N_2O and N_2 emissions and denitrifier community size and activity. *Applied and Environmental Microbiology* **76**:1870-1878.

Eddy, S.R., (1998) Profile hidden Markov models. *Bioinformatics Review* **14**:755-763.

Edgar, R.C.,(2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**:113.

Finlayson-Pitts, B.J. and Pitts, J.N. (2000) *Chemistry of the Upper and Lower Atmosphere* (Academic Press, San Diego).

Forster, P.V., et al. eds (2007) *Changes in Atmospheric Constituents and in Radiative Forcing* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA).

Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S.P., Asner, G.P., Cleveland, C., Green, P. Holland, E. Karl, D.M., Michaels, A.F., Porter, J.H. Townsend, and Vörösmarty, A. 2004. Nitrogen Cycles: Past, Present and Future. *Biogeochemistry* **70**:153-226.

Haltia, T., Brown, K., Tegoni, M., Cambillau, C., Saraste, M., Kimmo M. K., and Djinovic-Carugo, K. (2003) Crystal structure of nitrous oxide reductase from *Paracoccus denitrificans* at 1.6 A resolution *Biochemical Journal* **369**: 77-88.

Heikkilä, M.P., Honisch, U., Wunsch, P., and Zumft, W.G. (2001) Role of the Tat transport system in nitrous oxide reductase translocation and cytochrome cd_1 biosynthesis in *Pseudomonas stutzeri*. *Journal of Bacteriology* **183**: 1663-1671

Honisch, U. and Zumft, W.G. (2003) Operon structure and regulation of the *nos* gene region of *Pseudomonas stutzeri*, encoding an ABC-type ATPase for maturation of nitrous oxide reductase. *Journal of Bacteriology* **185**:1895-1902.

Lashof, D.A. and Ahuja, D.R. (1990) Relative contributions of greenhouse gas emissions to global warming. *Nature* **344**:529-531.

Liu, X., Gao C., Zhang A., Jin P., Wang L. and Feng L. (2008). The *nos* gene cluster from gram-positive bacterium *Geobacillus thermodenitrificans* NG80-2 and functional characterization of the recombinant NosZ. *FEMS Microbiology Letters* **289**: 46-52.

Machida, T., Nakazawa, T., Fujii, Y., Aoki, A., and Watanabe, O. (1995) Increase in the atmospheric nitrous oxide concentration during the last 250 years. *Geophysical Research Letters* **22**:2921-2924.

McSwiney, C.P. and Robertson, P.G. (2005) Nonlinear response of N_2O flux to incremental fertilizer addition in a continuous maize (*Zea mays* L.) cropping system. *Global Change Biology* **11**:1712-1719.

Paraskevopoulos, K., Antonyuk, S.V., Sawers, R.G., Eady, R.R., and Hasnain, S.S. (2006) Insight into Catalysis of Nitrous Oxide Reductase from High-resolution Structures of Resting and Inhibitor-bound Enzyme from *Achromobacter cycloclastes.* *Journal of Molecular Biology* **362**: 55-65.

Payne, W.J., Grant, M.A., Shapleigh, J., and Hoffman, P. (1982) Nitrogen oxide reduction in *Wolinella succinogenes* and *Campylobacter* species. *Journal of Bacteriology* **152**:915-918.

Pomowski, A., Zumft, W.G., Kroneck, P.M.H., and Einsle, O. (2011) N_2O binding at a [4Cu:2S] copper–sulphur cluster in nitrous oxide reductase. *Nature* **477**: 234-238.

Ravishankara, A.R., Daniel, J.S., and Portmann, R.W. (2009) Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. *Science* **326**:123-125.

Riester, J., Zumft, W.G., and Kroneck, P.M.H. (1989) Nitrous oxide reductase from *Pseudomonas stutzeri*. *European Journal of Biochemistry* **78**: 751-762.

Sanford, R.A., Cole, J.R., and Tiedje, J.M. (2002) Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68**:893-900.

Sanford, R.A., Wagner, D.D., Wu, Q., Chee-Sanford, J.C., Thomas, S.H., Cruz-Garcia, C., Rodriguez, G., Massol-Deya, A., Krishani, K.K., Ritalahti, K.M., Nissen, S., Konstantinidis, K.T., and Löffler, F. E. (2012) Unexpected non-denitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proceeding of the National Academy of Sciences of the United States of America* In Press.

Simon, J. (2002) Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS Microbiology Reviews* **26**: 285-309.

Simon, J., Einsle, O., Kroneck, P.M.H., and Zumft, W. (2004) The unprecedented *nos* gene cluster of *Wolinella succinogenes* encodes a novel respiratory electron transfer pathway to cytochrome *c* nitrous oxide reductase. *FEBS Letters* **569**:7-12.

Thomas, S.H., Padilla-Crespo, E., Jardine, P.M., Sanford, R.A., and Löffler, F.E. (2009) Diversity and distribution of *Anaeromyxobacter* strains in a uranium-contaminated subsurface environment with a nonuniform groundwater flow. *Applied and Environmental Microbiology* **75**:3679-3687.

Velasco, L., Mesa, S., Xu, C.A., Delgado, M.J., and Bedmar, E.J. (2004) Molecular characterization of *nosRZDFYLX* genes coding for denitrifying nitrous oxide reductase of *Bradyrhizobium japonicum*. *Antonie Leeuwenhoek International Journal* **85**:229-235.

Wallenstein, M.D., Myrold, D.D., Firestone, M., and Voytek, M. (2006) Environmental controls on denitrifying communities and denitrification rates: Insights from molecular methods. *Ecological Applications* **16**:2143-2152.

Wrage, N., Velthof, G.L., van Beusichem, M.L., and Oenema, O. (2001) Role of nitrifier denitrification in the production of nitrous oxide. *Soil and Biochemistry* **33**: 1723-1732.

Wu, Q., Sanford R. A. and Löffler F. E. (2006). Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. *Applied and Environmental Microbiology* **72**: 3608-3614.

Zumft, W.G. (1997) Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews* **61**:533-616.

Zumft, W.G. and Körner, H. (2007) Nitrous oxide reductases. *Biology of the nitrogen cycle*, eds Bothe, H., Ferguson, S.J., & Newton, W.E. (Elsevier, Amsterdam), pp 67-81.

CHAPTER 7

IMPLICATIONS AND FUTURE PERSPECTIVES.

7.1 Genomics and further characterization of organohalide respiration.

The current studies provided insights into respiration, coenzyme acquisition, and horizontal gene transfer (HGT) among OHRB genomes which may facilitate further investigations of organohalide respiration in geochemical cycling or contaminant mitigation. The findings presented here serve to advance two broad questions; first, which metabolic functions, reflected on the genomes, are shared across all OHRB, and second, which genomic features separate OHRB strains from related non-organohalide respirers? RDases and the corrinoid ABC transport cluster, *btuF-C-D*, were among the few genes found to be conserved across all OHRB genomes, while the latter is by no means unique to OHRB. Yet, as evidenced by the incomplete or missing corrinoid biosynthesis pathways on *Dehalococcoides mccartyi* and *Anaeromyxobacter dehalogenans* (chapter 5), the minimal set of genes required for organohalide respiration may not necessarily be found within a single organism. Metagenomic sequencing of organohalide-respiring microbial communities suggests individual OHRB strains need not harbor the full set genes essential to organohalide respiration. For example, catabolic genes involved in the conversion of carbon sources (e.g., lactate) to H_2, as well as the full set of *de novo* corrinoid biosynthesis genes, were harbored in non-organohalide-respiring members of chloroethene-degrading consortia (Hug et al., 2012). In addition, expression studies on the *Dehalococcoides*-containing KB-1 metagenome has shown that non-organohalide respirers in the KB-1 consortium along with putative bacteriophages appear to play an active role in responses to

chloroethenes (Hug et al., 2011; Waller et al,. 2012). As high-throughput sequences of microbial communities containing OHRB become increasingly available, insights into the OHRB genomic features elucidated in the current study may enable systems biology based characterization of organohalide-degrading consortia (Vietes et al., 2008; Vandenkoornhuyse et al., 2010; Gilbert and Dupont, 2011. The current study will assist such future metagenomic efforts by revealing the distinguishing features within key terminal reductases, such as class 1 RDase genes and *nosZ*, which may enhance single-gene environmental assays for organohalide respiration and N_2O reduction, respectively. As over 90% of class 2 sequences plus a small percentage of class 1 sequences (<3%) from the RDase superfamily were from organisms lacking confirmed activity towards organohalides, RDase genes may not always function in organohalide respiration (chapter 3). Yet, multiple sets of class 1 RDase genes and corrinoid salvage genes (e.g., *cbiZ*), appear to be features found only on OHRB genomes (chapters 3 and 5). Hence, the current study has advanced our understanding of genes essential to OHRB or abundant on OHRB genomes and delineated biomarkers distinguishing organohalide respiring consortia from microbial communities incapable of organohalide respiration.

Aside from functional characterization of organohalide-respiring organisms and communities, the current study motivates further inquiry into the evolution of genes and organisms in response to anthropogenic changes. RDase genes were shown to occur on mobile genomic elements detectable by established computational techniques (Chapter 4), while a customized method indicated occasional RDase transfer beyond the genus level (Chapter 3). Furthermore, mobile genomic elements such as plasmid pSZ77 and chromosomal genomic islands may have played a role in the divergence of the PCE-respiring *Geobacter lovleyi* from non-organohalide-

respiring *Geobacter* (chapter 4). Despite evidence of RDase HGT, it seems unlikely that non-organohalide respirers readily acquire functional RDases from OHRB. By contrast, corrinoid salvage or transport genes appear to be easily transferred between non-OHRB and OHRB, as evidenced by ancient HGT of *Chloroflexi cbiZ* from *Archaea* (Chapter 5). Besides elucidating the role of HGT in organohalide respiration, the current study may also motivate future investigations into convergent evolution or co-evolution among OHRB. For instance, the origin of PCE RDases in multiple evolutionary lineages of the RDase superfamily (chapter 3) may facilitate the adaptation of OHRB to anthropogenic chloroethenes, even when HGT is restricted. As another example, association between the pathways of organohalide respiration and corrinoid salvage apparent on the genomes of *Dehalococcoides mccartyi* (chapters 5) supports evidence that OHRB respiratory metabolism and corrinoid biosynthesis may have co-evolved (Maphosa et al., 2012). Hence, increased availability of sequencing data from OHRB promises to expand our current understanding of the relationships between ecology and evolution.

7.2 Implications, outlook, and recommendations.

The current comparative genomics-based studies provided opportunities to advance bioinformatics methodologies for groups of environmental microbes underrepresented in traditional, wet-lab experimental studies. Emerging sequencing technologies produce increasing amounts of environmental sequence data at diminishing cost; yet functional gene annotation is often inadequate or not readily applicable to bioremediation concerns (Desai et al., 2010; Metzger et al., 2011). Inference of RDase A protein domain structure, RDase A – RDase B interactions, and RDase gene evolutionary history in chapter 3 represents the type of bioinformatics analyses

critically needed for genome sequences relevant to bioremediation (Vandenkoornhuyse et al., 2010). Previous analyses of OHRB genomes identified putative RDases based upon vaguely defined homology with characterized RDases (chapters 2 and 3). In the current study, combined HMM and manual sequence curation delineated amino acid sequence features (domains and motifs) serving to discriminate RDases from unrelated iron-sulfur proteins (chapter 3), providing reproducible criteria for identification of putative RDases. This study went beyond simple counts of genomic RDases to integrate custom tools for detection of key functional features, such as Tat-type N-terminal signals, with HMMs and MSAs to divide the identified RDase superfamily into two functional classes (chapter 3). Similar HMM- and MSA-based methods enabled detection of an atypical subfamily of NosZ proteins in *Anaeromyxobacter* and *Desulfomonile*, two genera that were not expected to be capable of N_2O reduction (chapter 6). While HMMs are a well-established tool for inference of protein function and evolutionary history (Baldi et al., 1994; Eddy, 1998), the current study likely represents the first application of HMMs to terminal reductase genes and proteins of microbes critical to bioremediation. The methodologies employed here for RDases and NosZ thus provide a reference bioinformatics framework for investigators studying the classification of proteins exhibiting small numbers of highly-conserved functional residues (i.e., motifs) and multiple domains.

Key areas for further improvement in this study include adaptation of the bioinformatics methodology to high-throughput sequence annotation and investigation into electron donor metabolism of OHRB. First, the applicability of the HMMs, MSAs, and phylogenetic analyses to high-throughput pipelines may not have been fully explored. For instance, metagenomic datasets may be comprised of up to 7

million reads (Gilbert and Dupont, 2011) rendering the manual-curation-based methodologies of this study infeasible. At the same time, manual curation of RDases in chapter 3 provided essential validation steps towards integration of the RDase HMM into (meta)genome annotation pipelines. Thus, automating and scaling up these approaches will be important to keep up with the increasing volume of metagenomic data that become available. Similarly, the log fraction phylogenetic affiliation tool for determining Archaeal versus Bacterial gene origins in chapter 5 could be extended to whole genomes given further software development. A second possible shortcoming to the current study was the paucity of sequence analyses for genes involved in the utilization of organohalide respiratory electron donors, such as H_2 and acetate. *Geobacter lovleyi* strain SZ was shown in chapter 4 to harbor multiple genes encoding hydrogenases and pyruvate-ferredoxin oxidoreductases. Yet, the full set of genes which make H_2 available to *Dehalococcoides mccartyi* during organohalide respiration have thus far been deduced only through metagenomics (Hug et al., 2012; Maphosa et al., 2012). Hence, it appears that metagenomic (or metatranscriptomic) studies in association with curated OHRB genomes may be needed to provide a complete picture of OHRB metabolism. In summary, the bioinformatics methodologies employed in this study can be further optimized for automated annotation of genomes and metagenomes from organohalide-respiring communities.

7.3 References.

Baldi, P., Chauvin, Y., Hunkapiller, T., and McClure, M.A. (1994) Hidden Markov models of biological primary sequence information. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 1059-1063.

Desai, C., Pathak, H., and Madamwar, D. (2010) Advances in molecular and ‘‘-omics” technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites. *Bioresource Technology* 101: 1558-1569.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics Review* **14**: 755-763.

Friedberg, Iddo. (2007) Automated protein function prediction -- the genomic challenge. *Briefings in Bioinformatics* **7**: 225-242.

Gilbert, J.A. and Dupont, C.L. (2011) Microbial metagenomics: Beyond the genome. *Annual Review of Marine Science* **3**: 347–371.

Hug, L.A., Beiko, R.G., Rowe, A.R., Richardson, R.E., and Edwards, E.A. (2012) Comparative metagenomics of three *Dehalococcoides*-containing enrichment cultures: the role of the non-dechlorinating community. *BMC Genomics* **13**: 327.

Hug, L.A., Salehi, M., Nuin, P., Tillier, E.R., and Edwards, E.A. (2011) Design and verification of a pangenome microarray oligonucleotide probe set for *Dehalococcoides* spp. *Applied and Environmental Microbiology* 77: 5361–5369.

Maphosa, F., Lieten, S.H., Dinkla, I., Stams, A.J., Smidt, H., and Fennell, D.E. (2012) Ecogenomics of microbial communities in bioremediation of chlorinated contaminated sites. *Frontiers in Microbiology* **3**: 351.

Metzger, K.J., Klaper, R., and Thomas, M.A. (2011) Implications of informatics approaches in ecological research. *Ecological Informatics* **6**: 4-12.

Vandenkoornhuyse, P., Dufresne, A., Quaiser, A., Gouesbet, G., Binet, F., Francez, A-J., Mahé, S., Bormans, M., Lagadeuc, Y., and Couée, I. (2010) Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics. *Ecology Letters* **13**: 776-791.

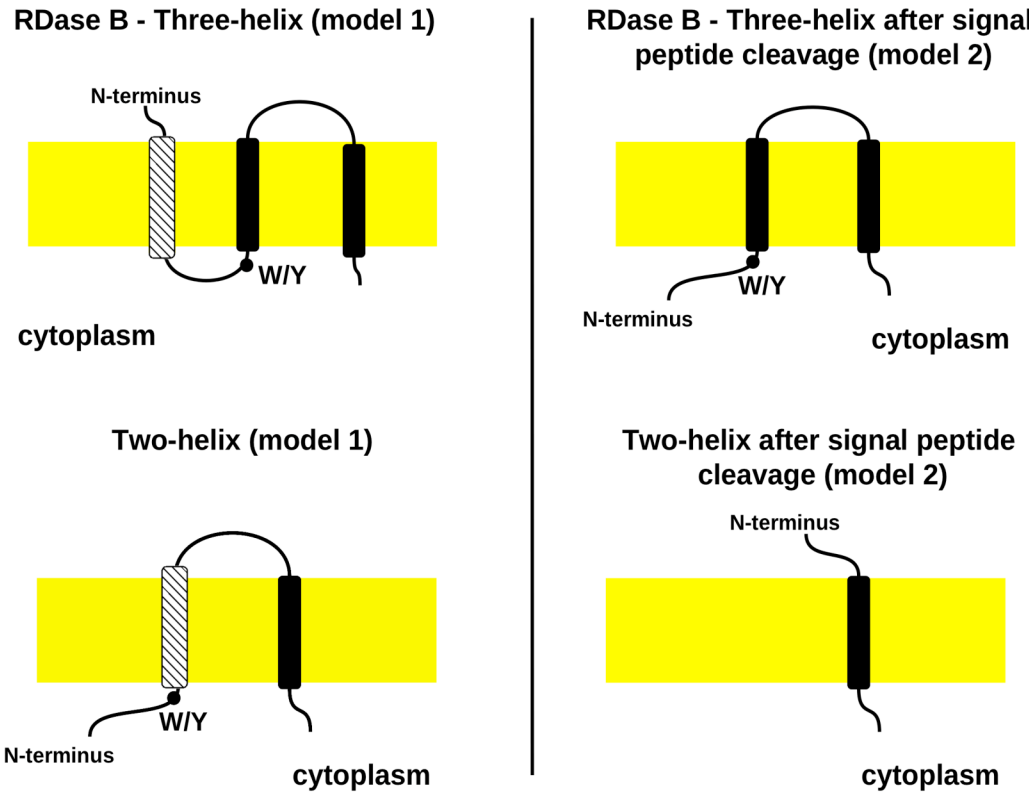
Vieites, J.M., Guazzaroni, M-E., Beloqui, A., Golyshin, P.N., and Ferrer, M. (2008) Metagenomics approaches in systems microbiology. *FEMS Microbiology Reviews* **33**: 236-255.

Waller, A.S., Hug, L.A., Mo, K., Radford, D.R., Maxwell, K.L., Edwards, E.A. (2012) Transcriptional analysis of a *Dehalococcoides*-containing microbial consortium reveals prophage activation. *Applied and Environmental Microbiology* 1178-1186

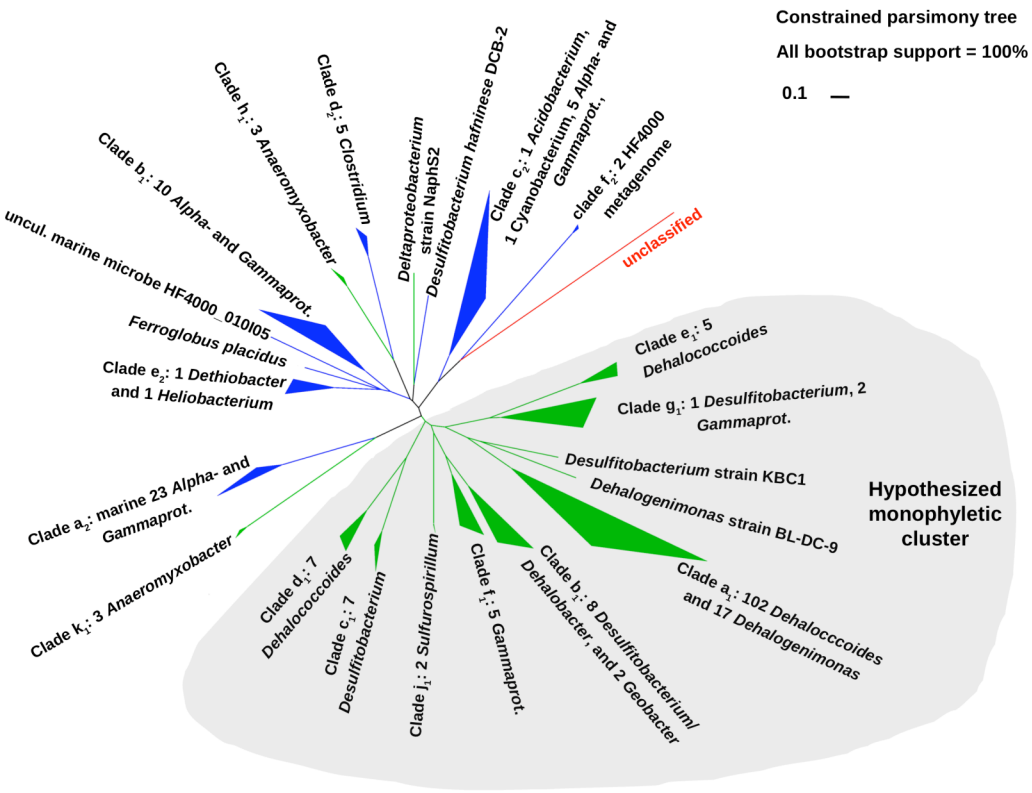
Appendix A, supplemental materials to Chapter 3.

```
Consensus                              RR F K
cbdb_A84_CbrA   -------------msnfhstvsRRdFmKa--lglagvglgaagaa-------tpvfhdldelisskpe---------vre
cbdbA1624_TcbA  -------------mkefhstlsRRdFmKs--lgvvgaglgtmsaa-------apvfhdldevtsstlg---------ink
DET0318_PceA    -------------mlnfhstltRKdFlKg--igmagaglgaasav-------apmfhdldelvastps---------trn
DhcVS1291_VcrA  -------------mskfhktisRRdFmKg--lglagagigavaas-------apvfhdidelvsseans--------tkd
DET0079_TceA    ------------msekyhstvtRRdFmKr--lglagagagalgaavlaennlphefkdvddllsagkalegdhank-vnn
BAV1_0847_BvcA  -------------mhnfhctisRRdFmKg--lglagagigaatsv-------mpnfhdldevisaasaetsslsgkslnn
AB194706_PrdA   -mdrekentldqkeekrsvgisRRnFfKasgiaagvaalglvtks-------qpvyagqesesaivnfavqevdqspynl
CAD28790_PceA   -----------------mgeinRRnFlKasmlgaaaaavasasav-------kgmvsplvadaadivapitetsefpykv
AAQ54585_cprA5  -----------------mnldRRsFlKaslvsvaavaaasaaaa-------ketfapltaeaaeiiapiretaefpyqv
AF403179_CprA   mtrsgkmeklkkneeskalsinRRnFlK---igaattamgligav-------kss-sqvaaaatdtl------------
AF115542_cprA   ---------menneqrqqtgmnRRsFlK---vgaaattmgvigai-------kap--akvanaaetm------------
AAC60788_PceA   ------------mekkkkpelsRRdFgKliigggaaatiapfgvp-------ganaaekeknaaeirqqfamtagspiiv

Consensus
cbdb_A84_CbrA   sp--------wwvkere-fedptveid---------wslkthfd---------y-------nlvhswvsketaqewqekq
cbdbA1624_TcbA  np--------wwvkerd-fknptvpid---------wskvtrqp-gv------f-------qglprptvadftkagvvgg
DET0318_PceA    lp--------wfvkere-hgdpttpid---------wdmiqrrpytwvrmdptl-------pvydnlksigapvsrwldw
DhcVS1291_VcrA  qp--------wyvkhre-hfdptitvd---------wdifdryd-gyqhkgvye-------gp-pdapftswgnrlqvrm
DET0079_TceA    hp--------wwvttrd-hedptcnid---------wslikrys-gwnnqgayflpedylsptytgrrhtivdskleiel
BAV1_0847_BvcA  fp--------wyvkerd-fenptidid---------wsilarnd-gynhqgayw-------gpvpeng----ddkrypdp
AB194706_PrdA   ppfanaenlkryelgknafyskelsmd-------------------------------------kfg--gnpwhiea
CAD28790_PceA   da-----kyqrynslkn-ffektfdpeanktpikfhyddvskit-gkkdtgkdl-------ptlnaerlgikgrpathte
AAQ54585_cprA5  dp-----kyqrlpaekl-aylrmfdpeenkgpikfhfddvskit-gkkdtgkdl-------pllnaeslgikgrpatlse
AF403179_CprA   ----------nyvpgrk-sqnsklhpe-------------------------------------hnygg-asvrfveh
AF115542_cprA   ----------nyvpgpt-narsklrpv-------------------------------------hdfag-akvrfven
AAC60788_PceA   nd-----kleryaevrtafthptsffk-------------------------------------pnykgevkpwflsa

Consensus                                  G
cbdb_A84_CbrA   aelireavanntpg-----ntl---kdmalanmglhyagsdlfnysqlsrpeystvildtfdvsgqidn-----------
cbdbA1624_TcbA  tstdletpemaltlydamakef-pgwtpGyagmgdvrttslcnaskfmmmgaw----pgnmemggkrinvigaimaaggs
DET0318_PceA    edkkaedeilyaka----redf-pgwepGldgfgdirttalthasemfsfgnf----ptrmnlggnmvdlvaavraaggy
DhcVS1291_VcrA  sgeeqkkrilaakk-----erf-pgwdgGlhgrgdqradalfyavtq----pfpgsgeeghglfqpypdqpg--------
DET0079_TceA    qgkkyrdsafiksgidwmkenidpdydpGelgygdrredaliyaatngshncwenplygryegsrpylsmrtmngi----
BAV1_0847_BvcA  adqcltlpekrdlylawakqqf-pdwepGinghgptrdealwfasstggigryripgtqqmmstmrld------------
AB194706_PrdA   ekyivkfikegvp--------------Gyslmdnafydaawasykg-------------------------------------
CAD28790_PceA   tsilfqtqhlgaml-----t-q-rhnetGwtgldealnagawavefdysgfnaagggpgsviplypinpmtneianepvm
AAQ54585_cprA5  tgaiffshhdgsvl-----p-l-rekemGwraldmalvvaswsveyhyngftapgsgpggviahypfnpmtnetgtepvf
AF403179_CprA   ndqwlgttkiigti-----knp-neadmGfnlaargilgdqakkg------------vnnfiakhpfgssiss-------
AF115542_cprA   ndewlgttkiiskv-----kkt-seadaGfmqavrglygpdpqrgff------------qfiakhpf-------------
AAC60788_PceA   ydekvrqiengeng-----pkm-kaknvGearagraleaagwtldinygniy-----pnrffmlwsgetmtntq------

Consensus                                        E                 G
cbdb_A84_CbrA   --------kygltrt----qlgipkwq--gtpEensamvaavlhflgstrvGyvsinennkkvwfspdka----------
cbdbA1624_TcbA  ptftpwlgpqldtttrpqdfgapvwq--gtpEenlktcrsafrffggsdvGaieidd-dvlkfihsqig----------
DET0318_PceA    lgstdsyagpkmvht--peemggtkyq--gtpEdnlrtlkagiryfggedvGaleldd-klkkliftvdqy---------
DhcVS1291_VcrA  -kfyarwglygpphdsappdgsvpkwe--gtpEdnflmlraaakyfgaggvGalnladpkckkliykkaqpmtlgkgtys
DET0079_TceA    -nglhefghadiktt------nypkwe--gtpEenllimrtaaryfgassvGaikitd-nvkkifyakvqpfcl-gpwyt
BAV1_0847_BvcA  -gstggwgyfnqppa-avwggkyprwe--gtpEentlmmrtvcqffgyssiGvmpits-ntkklffekqipfqfmagdpg
AB194706_PrdA   -tplfsweplgvsni--kraetvgkwe--atpEqnnryikkvaneygsgdtGvavlne----qwflsqdek---------
CAD28790_PceA   vpglynwdnidvesv----rqqgqqwkf-eskEeaskmvkkatrllgadlvGiapyde----rwtystwgrkil-kpckm
AAQ54585_cprA5  lagmyswdntkarer----reqgrqwkfesv-EeasrivkkaarflgadmaGiapydd----rwtfstwcrpnl-kpfkl
AF403179_CprA   -glmfiageeavtgq--pakeklpi----pdpEqmsqhikdaayfladevGigrmpe----fayyshkt----------
AF115542_cprA   -ggtiswarnliaaedvvdgdaeptktpipdpEqmsqhirdccyflradevGigkmpe----ygyythhv----------
AAC60788_PceA   -----lwapvgldrr--------ppdt--tdpveltnyvkfaarmagadlvGvarlnr----nwvyseavt----ipadv

Consensus
cbdb_A84_CbrA   -------------------nriiswgdveepvntpgilwpgnklgslvlpnkcnslisfvipqsgis---kyhhtalsra
cbdbA1624_TcbA  -------------------gkavvvedveeayett---------tkmviprkckwilmwsarqsleat--rrqagitesh
DET0318_PceA    -------------------gkalefgdveecietp---------kqvtipnkckyiflwtmrqpyewt--rrqsgrfega
DhcVS1291_VcrA  eiggpgmidakiypkvpdhavpinfkeadysyynd---------aewviptkcesiftftlpqpqel---nkrtggiaga
DET0079_TceA    itnmaeyieypv--pvdnyaipivfedipadqghysykrfggd-dkiavpnaldniftytimlpekrf--kyahsipmdp
BAV1_0847_BvcA  vfggtgnvqfdvplpkt--pvpivweevdkgyynd---------qkivipnkanwvltmtmplpedrf--krslgwslda
AB194706_PrdA   -------------------gkpyvfstehskptite--------eayyipktmnrvivmlapmnpnml--kyapttlsea
CAD28790_PceA   pngrtkylpwdlpkmlsg-ggvevfghakfepdwek----------yagfkpksvivfvleedyeai--rtspsvissa
AAQ54585_cprA5  pngrteyfltdpfklmk--gevevygstsveadwek----------yagftpksviamtfemdyeay--rtapsvlqga
AF403179_CprA   -------------------tvprtkmaaapveewt----------vpfeekhpyvivvmvdqhlettlastgydgisga
AF115542_cprA   -------------------sdtvglmskpveecvt----------pvtkiypnvivvmidqgietmwastgydgisga
AAC60788_PceA   pyeqslhkei---------ekpivfkdvplpietdd---------eliipntcenvivagiamnremm--qtapnsmaca
```

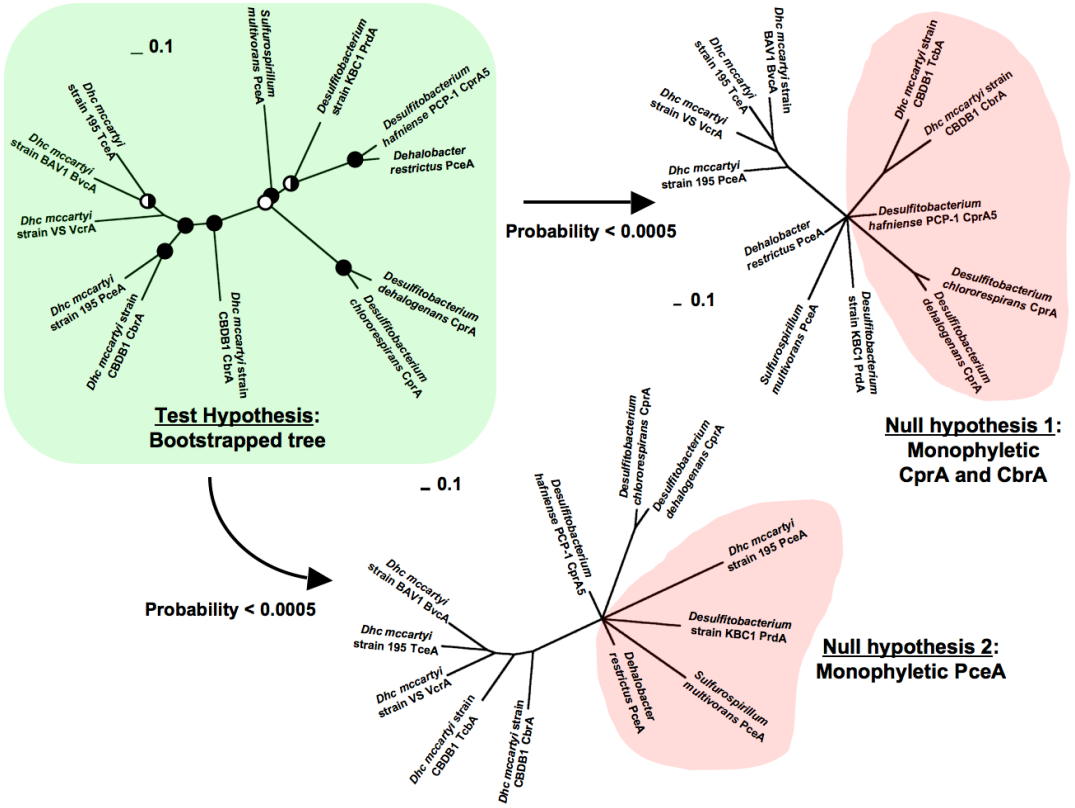
A.1 Figure S1. Characterized RDase MSA. Muscle alignment of 12 biochemically and transcriptionally characterized RDases with 80% consensus residues plotted above alignment.



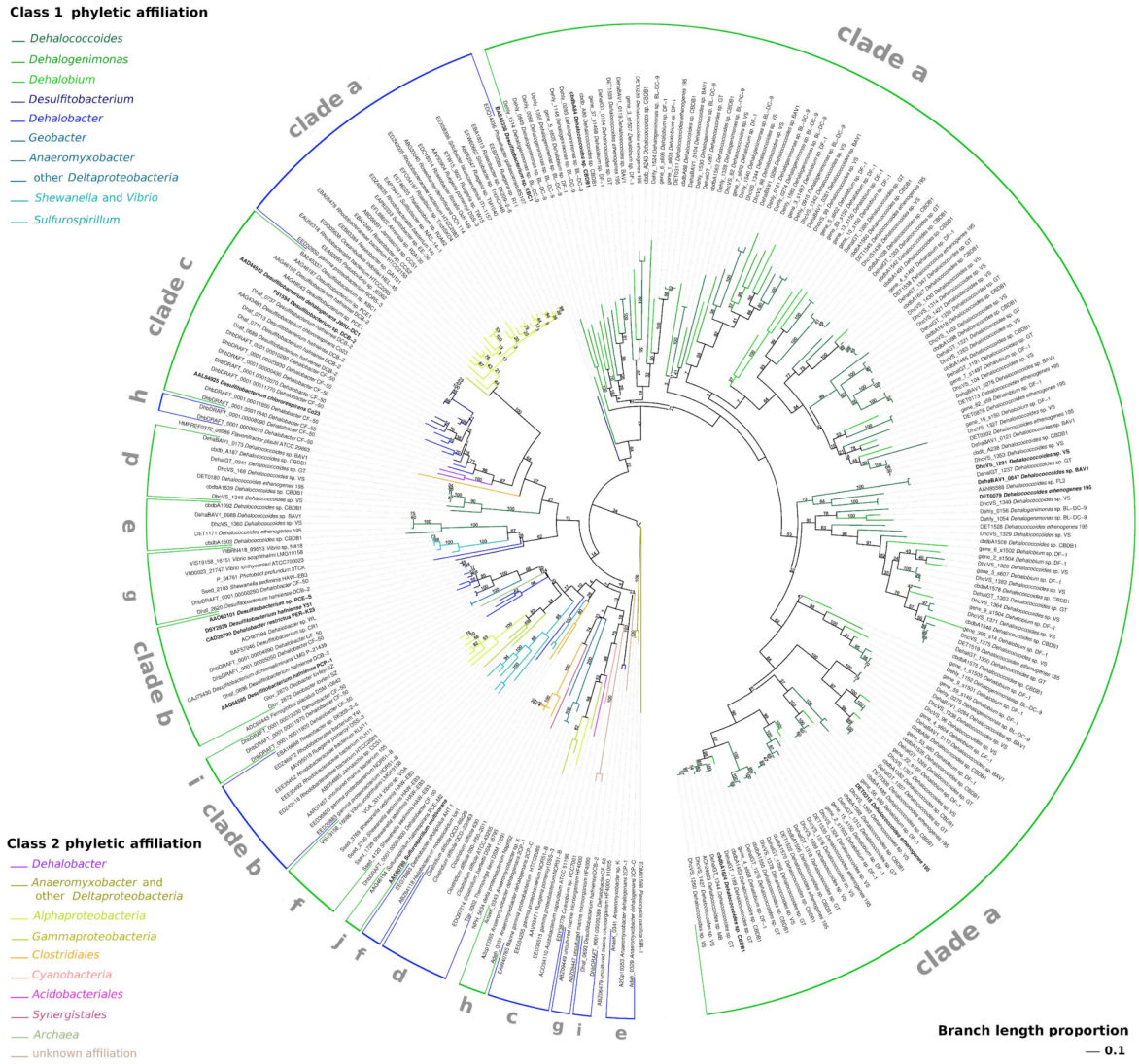
A.2 Figure S2. Membrane topologies of predicted RDase B subunit proteins. Transmembrane (TM) helices and topologies inferred from the complete protein (model 1), and the protein with the signal peptide cleaved of (model 2). Yellow band denotes the cell membrane, black bars denote TM helices, diagonal hatched bars denote TM helices lying within the Sec-type signal peptide region. W/Y denotes the aromatic-rich motif (**WY**x**W** in *Dehaloccoccoides mccartyi*) facing the cytoplasm in at least 164 out of 180 B proteins. Prediction of TM helices and membrane topologies inferred by the TMHMM server (www.cbs.dtu.dk/services/TMHMM) and signal peptide cleavage sites inferred by Eukaryotic model SignalP (www.cbs.dtu.dk/services/SignalP).



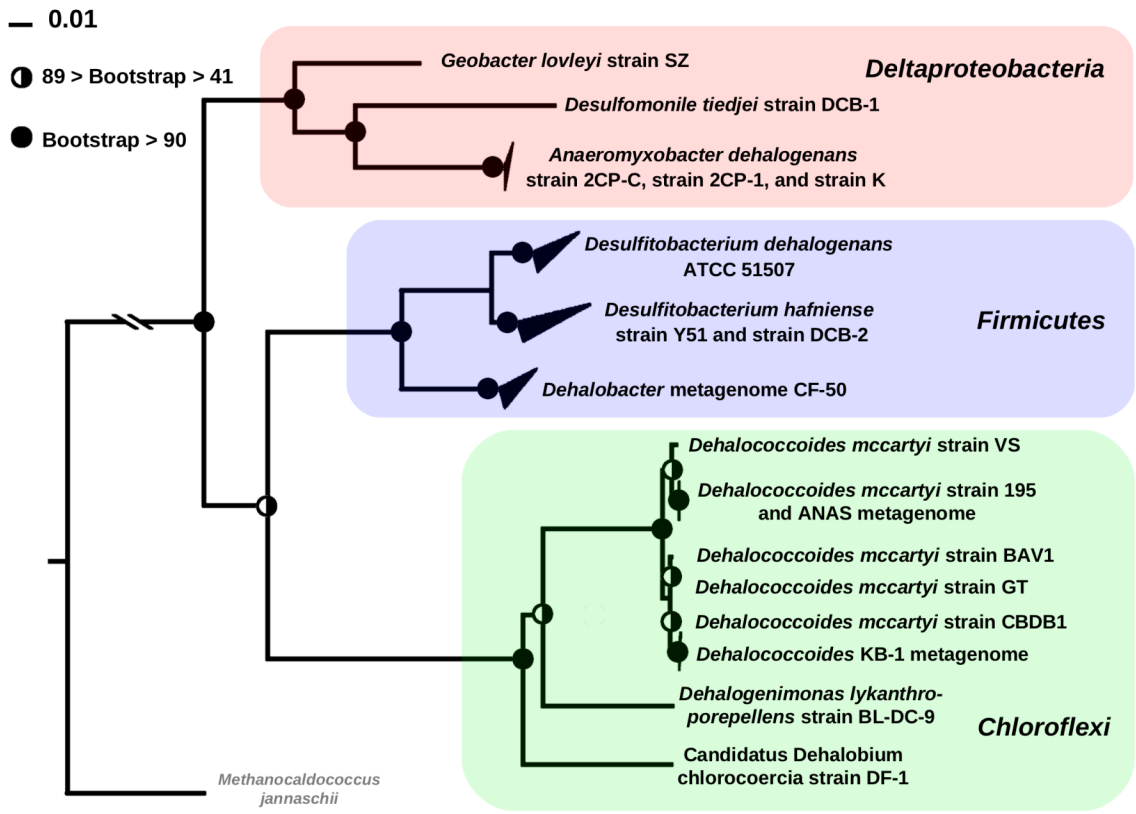
A.3 Figure S3. Tree testing null hypothesis of class 1 monophyly. Constrained parsimony 226-sequence tree with monophyly assumed for class 1 sequences possessing N-terminal Tat. Shared ancestry of Tat-containing class 1 sequences serves as a null hypothesis for comparison with the 226-sequence bootstrapped tree. The scale bar indicates a 10% change among aligned amino acid residues.



A.4 Figure S4. Polyphyly of substrate affinity among experimentally characterized RDases. Constrained parsimony trees assuming monophyly of RDases with chloroaromatic substrates (null hypothesis 1) and monophyly of PCE reductases (null hypothesis 2) exhibited probabilities < 0.00005 of representing the optimal topology represented by the bootstrapped tree.



A.5 Figure S5. Phylip bootstrapped tree of 283 RDase sequences. An unclassified iron-sulfur sequence served as the outgroup. Branches are colored according to predicted functional class and phyletic affiliation (genus or phylum) of host genome, where blues and dark greens represent class 1 sequences and warm colors represent class 2 sequences. Sequences representing biochemically characterized RDases are indicated in boldface. The scale bar indicates branch lengths corresponding to a 10% change in primary amino acid sequence. View at 300%.



A.6 Figure S6. 16S rRNA gene bootstrapped tree of closed genomes and characterized metagenomes of OHRB. Topological disagreement between *Chloroflexi* 16S tree and RDase phylogeny subtrees employed as a factor to infer possible HGT (criterion iii). The scale bar at the top left corresponds to a 1% change in nucleotides while half-closed and closed circles denote percent bootstrap support out of 1000 replicates. The non-OHRB *Archaeon*, *Methanocaldococcus jannachii* serves as outgroup.

A.7 Table S1. RDase A proteins with enzymological or transcriptional characterization of organohalide respiratory function. 12 of the 16 RDases shown were used to build the MSA (Fig. S1) and RDase HMM. Three characterized PceAs of *Desulfitobacterium* shared > 95% amino acid identity with the *Dehalobacter restrictus* strain PER-K23 PceA while the *Desulfitobacterium hafniense* strain DCB-2 CprA shared 99% identity with the *Desulfitobacterium dehalogenans* CprA and could be excluded from the RDase HMM without loss of sensitivity.

Organism	Symbol	Substrates	References
Dehalobacter restrictus *Desulfitobacterium hafniense* strain Y51, *Desulfitobacterium hafniense* strain TCE1, *Desulfitobacterium* sp. strain PCE-S	PceA*	PCE, TCE, and chloroethanes	(Maillard et al., 2003; Suyama et al., 2002; Maillard et al., 2005; Miller et al., 1998)
Desulfitobacterium hafniense Str. PCP-1	CprA5	3,5-dichlorophenol	(Thibodeau et al., 2004)
Desulfitobacterium chlororespirans strain Co23 *Desulfitobacterium dehalogenans* ATCC 51507 *Desulfitobacterium hafniense* strain DCB-2	CprA**	Chlorophenol and 3-chloro-4-hydroxy-phenylacetate	(Krasotkina et al., 2001; van den Pas et al., 1999)
Desulfitobacterium sp. KBC1	PrdA	PCE	(Tsukagoshi et al., 2005)
Sulfurospirillum multivorans	PceA	PCE and chloropropenes	(Neumann et al., 2002)
Dehaloccoccoides mccartyi strain 195	TceA	PCE, TCE and all DCEs	(Magnuson et al., 1998; Magnuson et al., 2000)
Dehaloccoccoides mccartyi strain 195	PceA	PCE	(Magnuson et al., 1998)
Dehalococcoides mccartyi strain CBDB1	CbrA	Tetra- and Trichlorobenzenes	(Adrian et al., 2007)
Dehalococcoides mccartyi strain CBDB1	CbrA	Trichlorobenzenes	(Wagner et al., 2009)
Dehalococcoides mccartyi strain VS	VcrA	*cis*-DCE and VC	(Müller et al., 2004)
Dehalococcoides mccartyi strain BAV1	BvcA	*cis*-DCE and VC	(Krajmalnik-Brown et al., 2004)

* *Desulfitobacterium hafniense* Str. Y51, *Desulfitobacterium hafniense* Str. TCE1 and *Desulfitobacterium* Str. PCE-S PceAs share > 96% identity with *Dehalobacter restrictus* PceA and are excluded from HMM.
** *Desulfitobacterium hafniense* strain DCB-2 CprA (Dhaf_0737) excluded from HMM due to high pairwise identity (91%) with *D. dehalogenans* CprA

A.8 Table S2. Publicly available closed genomes of organohalide-respiring bacteria (OHRB) as of July 2012.

Organism	Genome Accession	RefSeq-annotated RDases	Expert-curated RDases	References
Dehalococcoides mccartyi strain 195	CP000027	18	17	(Seshadri et al., 2006)
Dehalococcoides mccartyi strain CBDB1	AJ965256	32	32	(Kube et al., 2006)
Dehalococcoides mccartyi strain BAV1	CP000688	10	11	(McMurdie et al., 2009)
Dehalococcoides mccartyi strain VS	CP001827	37	36	(McMurdie et al., 2009)
Dehalococcoides mccartyi strain GT	CP001924	20	20	(Sung et al., 2006a)
Dehalogenimonas lykanthroporepellens strain BL-DC-9	CP002084	19	Not available	(Yan et al., 2009)
Desulfitobacterium hafniense strain Y51	AP008230	2	2*	(Nonaka et al., 2006)
Desulfitobacterium hafninese strain DCB-2	CP001336	7	7	(Kim et al., 2012)
Desulfitobacterium dehalogenans ATCC 51507	CP003348	6	Not available	(Utkin et al., 1994)
Desulfitobacterium dichloroeliminans strain LMG P-21439	AGJE01-000036	1	Not available	(Marzorati et al., 2007)
Geobacter lovleyi strain SZ	CP001089	2	2	(Wagner et al., 2012)
Anaeromyxobacter dehaloganans strain 2CP-C	CP000251	0†	2	(Sanford et al., 2002; Thomas et al., 2008)
Anaeromyxobacter dehaloganans strain 2CP-1	CP001359	0†	Not available	(Sanford et al., 2002)
Anaeromyxobacter dehaloganans strain K	CP001131	0†	Not available	(Sanford et al., 2002)
Desulfomonile tiedjei strain DCB-1	CP003360	1	Not available	(Mohn et al., 1992)

*Annotated chlorophenol reductive dehalogenase (DSY1155) not identified by RDase HMM in this study
† *Anaeromyxobacter dehalogenans* putative RDases (CprAs) annotated as "4Fe-4S ferredoxin"

A.9 Table S3. Class 1 sequences and key features.

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length (aa)	B-subunit present?	Function
Desulfitobacterium dichloroeliminans LMG P-21439	.	CAJ75430	Tat	551	yes	.
uncultured bacterium dca-cluster AM183919	.	CAJ75435	Tat	551	yes	
uncultured bacterium clone RDH1	.	BAF34631	Tat	486	yes	
uncultured bacterial clone RDH3 AB251616	.	BAF34632	Tat	462	yes	
uncultured *Dehalococcoides* strain clone KWI-A1	.	BAF34974	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-A2	.	BAF34975	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-C1	.	BAF34977	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-D1	.	BAF34978	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-E1	.	BAF34980	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-F1	.	BAF34981	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-F1	.	BAF34981	Tat	554	yes	
uncultured *Dehalococcoides* strain clone KWI-G1	.	BAF34982	Tat	540	yes	
uncultured *Dehalococcoides* strain clone KWI-LR1	.	BAF34984	Tat	554	yes	
Desulfitobacterium chlororespirans Co23 clone 1256 AF204275	.	AAG43483	Tat	456	yes	CprA
uncultured bacterium clone 8W-203 AB268344	.	BAF45397	Tat	519	yes	
uncultured bacterial clone 8W-601RRF2 AB268349	.	BAF45402	none	478	yes	
uncultured bacterial clone 8W-605RRF2 AB268350	.	BAF45404	Tat	484	yes	
uncultured bacterial clone 8W-607RRF2 AB268351	.	BAF45406	none	489	yes	
uncultured bacterium clone 8W-618RRF2	.	BAF45408	none	500	yes	
uncultured bacterial clone 8W-9A AB268353	.	BAF45410	none	497	yes	
uncultured bacterial clone 8W-9B AB268354	.	BAF45411	none	481	yes	
uncultured bacterial clone 8W-9E AB268355	.	BAF45413	none	445	yes	
uncultured bacterial clone 8W-9I AB268357	.	BAF45415	none	483	yes	
Desulfitobacterium strain CR1 clone AB301952	.	BAF57046	Tat	551	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0104	ABQ16695	Tat	515	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0112	ABQ16703	Tat	506	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0119	ABQ16710	none	480	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0121	ABQ16712	Tat	514	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0173	ABQ16764	Tat	455	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0276	ABQ16863	Tat	523	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0281	ABQ16868	Tat	470	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0284	ABQ16871	Tat	496	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0296	ABQ16882	Tat	512	yes	
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0847	ABQ17429	Tat	516	yes	BvcA
Dehalococcoides mccartyi strain BAV1	DehaBAV1_0988	ABQ17568	Tat	532	yes	
Dehalococcoides clone FtL-RDase-bvcA	.	ABV44386	Tat	505	yes	
uncult. *Dehalococcoides* Ft_Lewis clone FtL-RDase-vcrA	.	ABV44388	Tat	509	yes	
uncultured *Dehalococcoides* clone FtL-RDase-1638	.	ABV44390	none	489	yes	
uncultured *Dehalococcoides* clone FtL-RDase-1618	.	ABV44392	Tat	465	yes	
Shewanella sediminis HAW-EB3	Ssed_1729	ABV36340	Tat	472	yes	
Shewanella sediminis HAW-EB3	Ssed_2100	ABV36709	Tat	474	yes	
Shewanella sediminis HAW-EB3	Ssed_2103	ABV36712	Tat	507	yes	
Shewanella sediminis HAW-EB3	Ssed_3769	ABV38373	Tat	468	yes	
Shewanella sediminis HAW-EB3	Ssed_4120	ABV38724	Tat	489	yes	
Desulfitobacterium strain PCE1	.	AAG46187	Tat	447	yes	
Desulfitobacterium hafniense DCB-2 clone AY013365	.	AAG46192	Tat	447	yes	CprA
Sulfurospirillum halorespirans strain PCE-M2 clone AY013367	.	AAG46194	Tat	501	yes	
Dehalococcoides strain clone KS4KSRdA01	.	ABY28306	Tat	504	yes	
Dehalococcoides culture clone KS5KSRdA06	.	ABY28307	Tat	473	yes	
Dehalococcoides culture clone KS9KSRdA04	.	ABY28308	Tat	466	yes	
Dehalococcoides culture clone KS16KSRdA07	.	ABY28309	Tat	456	yes	
Dehalococcoides culture clone KS19KSRdA08	.	ABY28311	Tat	486	yes	
Dehalococcoides culture clone KS22KSRdA03	.	ABY28312	Tat	493	yes	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
Dehalococcoides culture clone KS31KSRdA02	.	ABY28315	none	485	yes	
Dehalococcoides culture clone KS40KSRdA08	.	ABY28316	none	483	yes	
Dehalococcoides culture clone KS45KSRdA11	.	ABY28317	Tat	474	yes	
Dehalococcoides culture clone RC3RCRdA5	.	ABY28320	Tat	501	yes	
Dehalococcoides culture clone RC4RCRdA7	.	ABY28321	none	465	yes	
Dehalococcoides culture clone RC11RCRdA14	.	ABY28322	Tat	483	yes	
Dehalococcoides culture clone RC24RCRdA02	.	ABY28325	Tat	491	yes	
Dehalococcoides culture clone RC26RCRdA03	.	ABY28326	Tat	488	yes	
Dehalococcoides culture clone RC33RCRdA04	.	ABY28328	Tat	466	yes	
Dehalococcoides strain clone RC41RCRdA11	.	ABY28329	none	459	yes	
Dehalococcoides culture clone RC56RCRdA01	.	ABY28331	?	510	yes	
Dehalococcoides culture clone RC63RCRdA16	.	ABY28332	Tat	490	yes	
Dehalococcoides culture clone RC71RCRdA15	.	ABY28334	Tat	472	yes	
Dehalococcoides culture clone RC73RCRdA13	.	ABY28335	Tat	460	yes	
Geobacter lovleyi strain SZ	Glov_2870	ACD96581	Tat	514	yes	
Geobacter lovleyi strain SZ	Glov_2872	ACD96583	Tat	514	yes	
Desulfitobacterium chlororespirans Co23 clone 1256 AAL84925	.	AF403179	Tat	450	yes	CprA
Dehalococcoides strain MB clone RdhA5	.	ACF24861	Tat	489	yes	
Dehalococcoides strain MB clone DceA1 DceB1	.	ACF24863	Tat	496	yes	
Anaeromyxobacter dehalogenans strain K	AnaeK_0343	ACG71585	Sec	484	yes	
Dehalobacter strain WL clone FJ010189	.	ACH87594	Tat	551	yes	
uncultured bacterial clone TUT2264 rdhA5 AB362912	.	BAG72156	Tat	504	yes	
uncultured bacterial clone TUT2264 rdhA306 AB362914	.	BAG72158	Tat	504	yes	
uncultured bacterial clone TUT2264 rdhA1	.	BAG72159	Tat	496	yes	
uncultured bacterial clone TUT2264 rdhA322 AB362920	.	BAG72164	Tat	504	yes	
uncultured bacterial clone TUT2264 rdhA2 AB362921	.	BAG72165	Tat	491	yes	
uncultured bacterial clone TUT2264 rdhA6 AB362923	.	BAG72167	Tat	499	yes	
uncultured bacterial clone TUT2264 rdhA330 AB362924	.	BAG72168	Tat	503	yes	
uncultured bacterial clone TUT2264 rdhA8 AB362925	.	BAG72169	Tat	466	yes	
Desulfitobacterium hafniense DCB-2	Dhaf_0689	ACL18753	Tat	463	yes	
Desulfitobacterium hafniense DCB-2	Dhaf_0696	ACL18760	Tat	550	yes	
Desulfitobacterium hafniense DCB-2	Dhaf_0711	ACL18775	Tat	488	yes	
Desulfitobacterium hafniense DCB-2	Dhaf_0713	ACL18777	Tat	458	yes	
Desulfitobacterium hafniense DCB-2	Dhaf_0737	ACL18801	Tat	445	yes	CprA
*Desulfitobacterium hafniense*_DCB-2	Dhaf_2620	ACL20646	none	345	yes	
Anaeromyxobacter dehalogenans strain 2CP-1	A2cp1_0355	ACL63714	Sec	474	yes	
uncultured bacterium clone LO-RD-2	.	CAR57926	Tat	551	yes	
uncultured bacterium clone LO-RD-10	.	CAR57927	Tat	551	yes	
uncultured bacterium clone LO-RD-12	.	CAR57928	Tat	551	yes	
uncultured bacterium clone UP-RD-4	.	CAR57930	Tat	551	yes	
uncultured bacterium clone UP-RD-3	.	CAR57931	Tat	551	yes	
uncultured bacterium clone UP-RD-6	.	CAR57932	Tat	551	yes	
uncultured bacterium clone UP-RD-6'	.	CAR57933	Tat	551	yes	
uncultured bacterium clone UP-RD-G	.	CAR57934	Tat	551	yes	
uncultured bacterium clone UP-RD-D	.	CAR57935	Tat	551	yes	
uncultured bacterium clone UP-RD-A	.	CAR57936	Tat	551	yes	
uncultured bacterium clone UP-RD-M	.	CAR57937	Tat	551	yes	
uncultured bacterial clone 1 AB499746	.	BAI47790	none	453	yes	
uncultured bacterial clone 2 AB499747	.	BAI47792	none	446	yes	
uncultured bacterial clone 3 AB499748	.	BAI47794	none	455	yes	
uncultured bacterial clone 4 AB499749	.	BAI47796	none	477	yes	
uncultured bacterial clone 5 AB499750	.	BAI47798	none	455	yes	
uncultured bacterial clone AB499751	.	BAI47800	none	481	yes	
uncultured bacterial clone 7 AB499752	.	BAI47802	none	478	yes	
uncultured bacterial clone 8 AB499753	.	BAI47804	none	442	yes	
uncultured bacterial clone AB499754	.	BAI47806	none	454	yes	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
uncultured bacterial clone 10 AB499755	.	BAI47808	none	482	yes	
uncultured bacterial clone 9 AB499756	.	BAI47810	none	476	yes	
uncultured bacterial clone 12 AB499757	.	BAI47812	none	450	yes	
uncultured bacterial clone 14 AB499759	.	BAI47816	none	448	yes	
uncultured bacterial clone 15 AB499760	.	BAI47818	none	453	yes	
uncultured bacterial clone 16 AB499761	.	BAI47820	none	490	yes	
uncultured bacterial clone 17 AB499762	.	BAI47822	none	485	yes	
uncultured bacterial clone 18 AB499763	.	BAI47824	none	464	yes	
uncultured bacterial clone 19 AB499764	.	BAI47826	none	488	yes	
uncultured bacterial clone 20 AB499765	.	BAI47828	none	469	yes	
uncultured bacterial clone 21 AB499766	.	BAI47830	none	453	yes	
uncultured bacterial clone 22 AB499767	.	BAI47832	none	488	yes	
uncultured bacterial clone 23 AB499768	.	BAI47834	none	459	yes	
uncultured bacterial clone 24 AB499769	.	BAI47836	none	479	yes	
uncultured bacterial clone 25 AB499770	.	BAI47838	none	455	yes	
uncultured bacterial clone 26 AB499771	.	BAI47840	none	456	yes	
uncultured bacterial clone 27 AB499772	.	BAI47842	none	500	yes	
uncultured bacterial clone 28 AB499773	.	BAI47844	none	450	yes	
uncultured bacterial clone 29 AB499774	.	BAI47846	none	472	yes	
uncultured bacterial clone 30 AB499775	.	BAI47848	none	459	yes	
uncultured bacterial clone 31 AB499776	.	BAI47850	none	455	yes	
uncultured bacterial clone 32 AB499777	.	BAI47852	none	459	yes	
Vibrio strain RC586	VOA_3314	EEZ01311	Tat	470	yes	
Dehalococcoides mccartyi strain VS	DhcVS_82	ACZ61255	Tat	474	yes	
Dehalococcoides mccartyi strain VS	DhcVS_88	ACZ61261	Tat	512	yes	
Dehalococcoides mccartyi strain VS	DhcVS_96	ACZ61269	Tat	496	yes	
Dehalococcoides mccartyi strain VS	DhcVS_99	ACZ61272	Tat	470	yes	
Dehalococcoides mccartyi strain VS	DhcVS_104	ACZ61277	Tat	523	yes	
Dehalococcoides mccartyi strain VS	DhcVS_169	ACZ61341	Tat	455	yes	
Dehalococcoides mccartyi strain VS	DhcVS1260	ACZ62362	Tat	496	yes	
Dehalococcoides mccartyi strain VS	DhcVS1263	ACZ62364	Tat	497	yes	
Dehalococcoides mccartyi strain VS	DhcVS1291	ACZ62391	Tat	519	yes	VcrA
Dehalococcoides mccartyi strain VS	DhcVS1314	ACZ62413	Tat	481	yes	
Dehalococcoides mccartyi strain VS	DhcVS1316	ACZ62415	Tat	494	yes	
Dehalococcoides mccartyi strain VS	DhcVS1320	ACZ62419	Tat	509	yes	
Dehalococcoides mccartyi strain VS	DhcVS1324	ACZ62423	Tat	493	yes	
Dehalococcoides mccartyi strain VS	DhcVS1327	ACZ62426	Tat	499	yes	
Dehalococcoides mccartyi strain VS	DhcVS1329	ACZ62428	Tat	473	yes	
Dehalococcoides mccartyi strain VS	DhcVS1340	ACZ62439	Tat	480	yes	
Dehalococcoides mccartyi strain VS	DhcVS1342	ACZ62441	Tat	473	yes	
Dehalococcoides mccartyi strain VS	DhcVS_1344	ACZ62443	Tat	488	yes	
Dehalococcoides mccartyi strain VS	DhcVS_1349	ACZ62448	Tat	461	yes	
Dehalococcoides mccartyi strain VS	DhcVS1353	ACZ62452	Tat	514	yes	
Dehalococcoides mccartyi strain VS	DhcVS_1360	ACZ62459	Tat	532	yes	
Dehalococcoides mccartyi strain VS	DhcVS1364	ACZ62463	Tat	518	yes	
Dehalococcoides mccartyi strain VS	DhcVS1371	ACZ62470	Tat	505	yes	
Dehalococcoides mccartyi strain VS	DhcVS1375	ACZ62474	Tat	505	yes	
Dehalococcoides mccartyi strain VS	DhcVS1378	ACZ62477	Tat	508	yes	
Dehalococcoides mccartyi strain VS	DhcVS1383	ACZ62482	Tat	505	yes	
Dehalococcoides mccartyi strain VS	DhcVS1387	ACZ62486	Tat	491	yes	
Dehalococcoides mccartyi strain VS	DhcVS1393	ACZ62492	none	495	yes	
Dehalococcoides mccartyi strain VS	DhcVS1399	ACZ62498	Tat	494	yes	
Dehalococcoides mccartyi strain VS	DhcVS1402	ACZ62501	Tat	499	yes	
Dehalococcoides mccartyi strain VS	DhcVS1421	ACZ62520	Tat	475	yes	
Dehalococcoides mccartyi strain VS	DhcVS1427	ACZ62526	Tat	496	yes	
Dehalococcoides mccartyi strain VS	DhcVS1430	ACZ62529	Tat	492	yes	
Dehalococcoides mccartyi strain VS	DhcVS1436	ACZ62535	Tat	523	yes	
Dehalococcoides strain FL2 tceAB clone	.	AAN85588	Tat	554	yes	TceA
uncultured bacterium clone YK-TCE1	.	AAN85590	Tat	554	yes	
Uncultured bacterium clone PM-VC1	.	AAN85592	Tat	554	yes	
uncultured bacterium clone RC-VC2	.	AAN85594	Tat	554	yes	
uncultured marine bacterium clone rdh63A AF462232	.	AAO15649	none	438	yes	
uncultured bacterial clone YG1 AB526224	.	BAI70447	none	472	yes	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
uncultured bacterial clone YG2 AB526225	.	BAI70448	none	469	yes	
uncultured bacterial clone YG3 AB526226	.	BAI70449	none	470	yes	
uncultured bacterial clone YG6 AB526229	.	BAI70452	none	483	yes	
uncultured bacterial clone YG8 AB526231	.	BAI70454	none	473	yes	
uncultured bacterial clone YG9 AB526232	.	BAI70455	none	450	yes	
uncultured bacterial clone YG10 AB526233	.	BAI70456	none	481	yes	
uncultured bacterial clone YG11 AB526234	.	BAI70457	none	465	yes	
Dehalococcoides mccartyi strain GT	DehalGT_0124	ADC73492	Tat	482	yes	
Dehalococcoides mccartyi strain GT	DehalGT_0241	ADC73608	Tat	455	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1189	ADC74548	Tat	495	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1191	ADC74550	Tat	497	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1237	ADC74596	Tat	519	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1269	ADC74627	Tat	495	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1276	ADC74632	Tat	508	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1285	ADC74641	Tat	513	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1287	ADC74643	Tat	465	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1295	ADC74651	Tat	507	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1300	ADC74655	Tat	505	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1303	ADC74658	Tat	505	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1307	ADC74662	Tat	491	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1312	ADC74667	Tat	495	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1318	ADC74673	Tat	494	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1321	ADC74676	Tat	499	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1338	ADC74693	Tat	475	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1344	ADC74699	Tat	495	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1347	ADC74702	Tat	492	yes	
Dehalococcoides mccartyi strain GT	DehalGT_1353	ADC74708	Tat	500	yes	
Desulfitobacterium strain PCE-S clone	.	AAO60101	Tat	551	yes	PceA
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0156	ADJ25486	Tat	467	yes	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0275	ADJ25599	Tat	493	yes	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1054	ADJ26355	Tat	473	yes	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1152	ADJ26448	Tat	458	yes	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1524	ADJ26807	Tat	482	yes	
Sulfurospirillum multivorans clone AF022812	.	AAC60788	Tat	501	yes	PceA
Dehalococcoides mccartyi strain CBDB1 clone rdhA1 rdhB1	.	AAR24296	none	497	yes	
Dehalococcoides mccartyi strain CBDB1 clone RdhA2 rdhB2	.	AAR24298	Tat	484	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA3 rdhB3	.	AAR24300	none	482	yes	
Dehalococcoides mccartyi strain CBDB1 clone RdhA4 rdhB4	.	AAR24302	none	459	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA5 and rdhB5	.	AAR24304	none	476	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA6 rdhB6	.	AAR24306	none	471	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA7 rdhB7	.	AAR24308	none	449	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA8 rdhB8	.	AAR24310	none	479	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA10 rdhB10	.	AAR24313	none	478	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA11 rdhB11	.	AAR24315	none	479	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA12 rdhB12	.	AAR24317	none	489	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA13 rdhB13	.	AAR24319	none	489	yes	
Dehalococcoides mccartyi strain CBDB1 clone rdhA14 rdhB14	.	AAR24321	none	489	yes	
Dehalococcoides mccartyi strain FL2 clone rdhA1 rdhB1	.	AAR24325	none	459	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA2 RdhB2	.	AAR24327	none	483	yes	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
Dehalococcoides mccartyi strain FL2 clone rdhA3 rdhB3	.	AAR24329	none	481	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA4 RdhB4	.	AAR24331	none	476	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA5 RdhB5	.	AAR24333	none	497	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA6 RdhB6	.	AAR24335	none	484	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA7 RdhB7	.	AAR24337	none	482	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA8 RdhB8	.	AAR24339	none	479	yes	
Dehalococcoides mccartyi strain FL2 clone RdhA11 RdhB11	.	AAR24345	none	484	yes	
Desulfitobacterium hafniense PCP-1 clone AY349165	.	AAQ54585	Tat	548	yes	CprA
Dehalococcoides mccartyi strain BAV1 clone RdhA1 RdhB1	.	AAT48548	none	463	yes	
Dehalococcoides mccartyi strain BAV1 clone RdhA2 RdhB2	.	AAT48550	none	486	yes	
Dehalococcoides mccartyi strain BAV1	.	AAT48552	none	478	yes	
Dehalococcoides mccartyi strain BAV1 clone RdhA4 RdhB4	.	AAT48554	none	472	yes	
Dehalococcoides mccartyi strain BAV1 clone RdhA5 RdhB5	.	AAT48556	none	479	yes	
Dehalococcoides mccartyi strain BAV1 clone RdhA6 RdhB6	.	AAT48558	none	496	yes	
Dehalococcoides mccartyi strain BAV1 clone RdhA7 RdhB7	.	AAT48560	none	507	yes	
Dehalobacter restrictus PER-K23 clone AJ439607	.	CAD28790	Tat	551	yes	PceA
Desulfitobacterium dehalogenans ATCC 51507	Desde_0607	AFL99058	Tat	447	yes	CprA
Desulfitobacterium dehalogenans ATCC 51507	Desde_0041	AFL98536	Tat	463	yes	
Desulfitobacterium dehalogenans ATCC 51507	Desde_0803	AFL99249	Tat	451	yes	
Desulfitobacterium dehalogenans ATCC 51507	Desde_4095	AFM02356	Tat	468	yes	
Dehalococcoides mccartyi strain 195	DET0079	AAW39060	Tat	554	yes	TceA
Dehalococcoides mccartyi strain 195	DET1545	AAW39214	Tat	500	yes	
Dehalococcoides mccartyi strain 195	DET1559	AAW39215	Tat	482	yes	
Dehalococcoides mccartyi strain 195	DET1538	AAW39229	Tat	492	yes	
Dehalococcoides mccartyi strain 195	DET1535	AAW39240	Tat	494	yes	
Dehalococcoides mccartyi strain 195	DET1528	AAW39256	Tat	469	yes	
Dehalococcoides mccartyi strain 195	DET1522	AAW39262	Tat	507	yes	
Dehalococcoides mccartyi strain 195	DET1519	AAW39273	Tat	505	yes	
Dehalococcoides mccartyi strain 195	DET1171	AAW39605	Tat	532	yes	
Dehalococcoides mccartyi strain 195	DET0876	ACZ62452	Tat	510	yes	
Dehalococcoides mccartyi strain 195	DET0318	AAW40342	Tat	495	yes	PceA
Dehalococcoides mccartyi strain 195	DET0311	AAW40356	Tat	515	yes	
Dehalococcoides mccartyi strain 195	DET0306	AAW40361	Tat	505	yes	
Dehalococcoides mccartyi strain 195	DET0302	AAW40373	Tat	514	yes	
Dehalococcoides mccartyi strain 195	DET0235	AAW40468	Tat	490	yes	
Dehalococcoides mccartyi strain 195	DET0180	AAW40575	Tat	455	yes	
Dehalococcoides mccartyi strain 195	DET0173	AAW40589	Tat	510	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A80	CAI82340	Tat	496	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A84	CAI82345	Tat	488	yes	CbrA
Dehalococcoides mccartyi strain CBDB1	cbdb_A88	CAI82350	Tat	515	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A96	CAI82358	Tat	506	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A187	CAI82436	Tat	455	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A238	CAI82474	Tat	514	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A243	CAI82479	Tat	491	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1092	CAI83191	Tat	532	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1453	CAI83480	Tat	495	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1455	CAI83482	Tat	497	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1491	CAI83513	Tat	482	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1495	CAI83519	Tat	505	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1503	CAI83526	Tat	526	yes	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
Dehalococcoides mccartyi strain CBDB1	cbdbA1508	CAI83531	Tat	463	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1535	CAI83558	Tat	495	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1539	CAI83563	none	460	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1542	CAI83566	Tat	487	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1546	CAI83570	Tat	505	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1550	CAI83574	Tat	508	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1560	CAI83583	Tat	513	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1563	CAI83586	Tat	465	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1570	CAI83594	Tat	507	yes	
Dehalococcoides mccartyi strain CBDB1	cbdb_A1575	CAI83599	Tat	505	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1578	CAI83602	Tat	505	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1582	CAI83606	Tat	491	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1588	CAI83612	none	495	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1595	CAI83618	Tat	494	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1598	CAI83621	Tat	499	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1618	CAI83638	Tat	475	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1624	CAI83644	Tat	495	yes	TcbA
Dehalococcoides mccartyi strain CBDB1	cbdbA1627	CAI83647	Tat	492	yes	
Dehalococcoides mccartyi strain CBDB1	cbdbA1638	CAI83653	Tat	500	yes	
Dehalococcoides strain FL2 clone RdhA12 RdhB12	.	AAZ91690	none	478	yes	
Dehalococcoides strain FL2 clone RdhA13 RdhB13	.	AAZ91692	none	489	yes	
Desulfitobacterium strain KBC1 clone AB194705	.	BAE45337	Tat	447	yes	
Desulfitobacterium strain KBC1 clone AB194706	.	BAE45338	Tat	463	yes	PrdA
Dehalococcoides KB1 clone KB1rdhAB1	.	ABA64523	none	459	yes	
Dehalococcoides KB1 culture clone KB1RdhAB2	.	ABA64525	none	483	yes	
Dehalococcoides KB1 clone KB1RdhAB3	.	ABA64527	none	482	yes	
Dehalococcoides KB1 clone KB1rdhAB4	.	ABA64529	none	476	yes	
Dehalococcoides KB1 culture clone KB1RdhA5	.	ABA64531	none	484	yes	
Dehalococcoides KB1 clone KB1RdhAB6	.	ABA64533	none	500	yes	
Dehalococcoides KB1 culture clone RdhA7 RdhB7	.	ABA64535	none	489	yes	
Dehalococcoides KB1 clone KB1RdhAB8	.	ABA64537	none	479	yes	
Dehalococcoides KB1 clone KB1rdhAB9	.	ABA64539	none	479	yes	
Dehalococcoides KB1 clone KBRdhAB10	.	ABA64541	none	481	yes	
Dehalococcoides KB1 culture clone KB1RdhAB11	.	ABA64543	none	479	yes	
Dehalococcoides KB1 culture clone KB1RdhAB12	.	ABA64545	none	474	yes	
Dehalococcoides KB1 clone KBRdhAB13	.	ABA64547	none	495	yes	
Dehalococcoides KB1 clone KB1RdhAB14	.	ABA64549	none	503	yes	
Dehalococcoides isolate FMC-TCE	.	ABB89703	Tat	554	yes	
Dehalococcoides isolate SFR-cis-DCE	.	ABB89705	Tat	554	yes	
Dehalococcoides isolate Owls-PCE	.	ABB89709	none	526	yes	
Dehalococcoides isolate LH-PCE	.	ABB89711	none	526	yes	
Dehalococcoides isolate BRS-PCE	.	ABB89715	Tat	554	yes	
Anaeromyxobacter dehalogenans strain 2CP-C	Adeh_0331	ABC80107	Sec	491	yes	
Desulfitobacterium hafniense strain Y51	DSY2839	BAE84628	Tat	551	yes	
Photobacterium profundum strain 3TCK	P_04761	EAS45659	Tat	506	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00000290	.	Tat	513	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00000430	.	Tat	440	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00000800	.	Tat	507	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00004990	.	Tat	551	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00005050	.	Tat	551	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00011770	.	Tat	462	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00011850	.	Tat	470	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00011920	.	Tat	429	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00011970	.	Tat	434	yes	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00012030	.	Tat	445	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00012070	.	Tat	469	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00023930	.	Tat	448	yes	
Vibrio scophthalmi LMG19158	VIS19158_16151	.	Tat	516	yes	
Desulfitobacterium hafniense DCB-2 clone 3185	CPRA_DESHD	AAL87763	Tat	446	yes	CprA
Candidatus Dehalobium chlorocoercia strain DF-1	gene_82 scaffold_0599	.	none	501	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_5 scaffold_0602	.	none	482	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_7 scaffold_0602	.	none	473	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_4 scaffold_0604	.	none	416	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_5 scaffold_0605	.	none	454	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_6 scaffold_0606	.	none	486	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_3 scaffold_0607	.	Tat	510	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_4 scaffold_0608	.	none	483	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_53 scaffold_0608	.	none	484	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_56 scaffold_0608	.	none	480	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_4 scaffold_1491	.	none	479	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_59 scaffold_1491	.	Tat	489	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_3 scaffold_1497	.	none	458	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_7 scaffold_1497	.	none	494	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_37 scaffold_1498	.	none	475	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_395 scaffold_1498	.	Tat	505	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_2 scaffold_1503	.	Tat	497	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_9 scaffold_1501	.	none	474	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_6 scaffold_1502	.	none	483	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_83 scaffold_1503	.	Tat	491	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_2 scaffold_1504	.	Tat	525	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_8 scaffold_1504	.	Tat	495	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_10 scaffold_1504	.	none	440	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_22 scaffold_1504	.	none	479	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_13 scaffold_1506	.	none	478	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_16 scaffold_1506	.	none	486	yes	
Candidatus Dehalobium chlorocoercia strain DF-1	gene_3 scaffold_1507	.	none	501	yes	
Desulfitobacterium strain Viet1 clone AF259791	.	AAG49544	Tat	447	no	
uncultured bacterial clone RDH4 AB251617	.	BAF34633	Tat	457	no	
uncultured bacterial clone RDH6 AB251618	.	BAF34634	none	492	no	

A.9 Table S3. (continued)

Species and Source Description	Locus Tag (if available)	Genbank Accession	Signal type	A-subunit length(aa)	B-subunit present?	Function
uncultured *Dehalococcoides* strain clone KWI-F1	.	BAF34981	Tat	554	no	
Desulfitobacterium strain PCE1 clone AF259790	.	AAG49543	Tat	447	no	
Desulfitobacterium hafniense TCE1 clone AY013362	.	AAG46189	Tat	310	no	
Desulfitobacterium hafniense PCP-1 clone AY013364	.	AAG46190	Tat	316	no	
Desulfitobacterium hafniense DCB-2 clone AY013366	.	AAG46193	Tat	316	no	
Dehalococcoides culture clone KS24KSRdA09	.	ABY28313	Tat	468	no	
uncultured bacterial clone TUT2264 rdhA30 AB362911	.	BAG72155	none	485	no	
uncultured bacterial clone TUT2264 rdhA303 AB362913	.	BAG72157	Tat	488	no	
uncultured bacterial clone TUT2264 rdhA7 AB362916	.	BAG72160	Tat	488	no	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00000380	.	Tat	464	no	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00010290	.	Tat	460	no	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00011840	.	Tat	460	no	
Dehalococcoides mccartyi strain VS	DhcVS1336	ACZ62435	Tat	496	no	
Dehalococcoides mccartyi strain VS	DhcVS_1347	ACZ62446	none	429	no	
uncultured bacterial clone YG4 AB526227	.	BAI70450	none	469	no	
uncultured bacterial clone YG5 AB526228	.	BAI70451	none	475	no	
uncultured bacterial clone YG7 AB526230	.	BAI70453	none	448	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0068	ADJ25402	Tat	460	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0121	ADJ25452	Tat	469	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0274	ADJ25598	Tat	472	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0283	ADJ25606	Tat	476	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0849	ADJ26151	Tat	475	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_0910	ADJ26212	Tat	464	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1148	ADJ26444	Tat	462	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1328	ADJ26618	Tat	476	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1355	ADJ26644	Tat	474	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1514	ADJ26797	Tat	455	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1523	ADJ26806	Tat	340	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1530	ADJ26812	Tat	473	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1540	ADJ26821	Tat	465	no	
Dehalogenimonas lykanthroporepellens BL-DC-9	Dehly_1582	ADJ26861	Tat	452	no	
Deltaproteobacterium strain NaphS2	.	EFK11122	Sec	478	no	
Vibrio ichthyoenteri ATCC700023	VII00023_21747	.	Tat	504	no	
Vibrio scophthalmi LMG19158	VIS19158_16086	.	Tat	474	no	
Vibrio sp. strain N418	VIBRN418_09513	.	Tat	504	no	
Candidatus Dehalobium chlorocoercia strain DF-1	.	.	Tat	482	no	
Candidatus Dehalobium chlorocoercia strain DF-1	.	.	Tat	465	no	
Candidatus Dehalobium chlorocoercia strain DF-1	.	.	Tat	466	no	
Desulfitobacterium dichloroeliminans LMG P-21439	.	CAJ75430	Tat	551	yes	
uncultured bacterium dca-cluster AM183919	.	CAJ75435	Tat	551	yes	
Desulfosporosinus orientis DSM 765	Desor_3897	AET69295	Tat	454	yes	
Anaeromyxobacter dehalogenans 2CP-1	A2Cp1_0353	ACL63712	Sec	640	yes	
Anaeromyxobacter dehalogenans 2CP-C	Adeh_0329	ABC80105	Sec	640	yes	
Anaeromyxobacter dehalogenans strain K	AnaeK_0341	ACG71583	Sec	640	yes	
Desulfomonile tiedjei DSM 6799 (DCB-1)	Desti_0783	AFM23508	Sec	559	yes	
Desulfomonile tiedjei DSM 6799 (DCB-1)	Desti_0785	AFM23510	Sec	351	no	
Desulfomonile tiedjei DSM 6799 (DCB-1)	Desti_1412	AFM24124	Tat	471	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00008070	.	none	434	yes	
Dehalobacter metagenome CF-50	DhbDRAFT_0001.00008090	.	none	439	yes	

A.10 Table S4. Class 2 sequences and key features.

Species / Source Description	Available Locus Tag	Genbank Accession	Signal type	Length (aa)
Roseobacter denitrificans OCh 114	RD1_4098	ABG33540	none	1070
Rhodobacterales bacterium HTCC2255	OM2255_0919	EAU52318	none	1052
marine *Gamma proteobacterium* HTCC2080	MGP2080_13613	EAW40765	none	340
Clostridium difficile strain 630	CD1958	CAJ68833	none	345
Rhodobacteraceae bacterium HTCC2150	.	EBA05478	none	1064
Roseobacter sp. strain CCS2	.	EBA12491	none	1074
Roseobacter sp. strain SK209-2-6	RSK20926_11359	EBA18313	none	1070
Roseobacter sp. strain SK209-2-6	RSK20926_02649	EBA16668	none	402
Phaeobacter gallaeciensis BS107	RGBS107_03023	EDQ14036	none	1052
Oceanibulbus indolifex HEL-45	.	EDQ05638	none	1068
Clostridium bartlettii DSM16795	CLOBAR_00966	EDQ97214	none	354
uncultured marine microorganism HF4000_010I05	ALOHA_HF4000010I05ctg1g44	ABZ06479	none	427
uncultured marine microorganism HF4000_APKG8C21	ALOHA_HF4000APKG8C21ctg1g35	.	none	416
uncultured marine microorganism HF4000_APKG8C21	ALOHA_HF4000APKG8C21ctg1g37	.	none	418
Heliobacterium modesticaldum Ice1	HM1_1546	ABZ84118	none	405
Cyanobium sp. strain PCC7001	CPCC7001_1657	EDY38778	none	452
Desulfitobacterium hafniense DCB-2	Dhaf_0693	ACL18757	none	352
Desulfitobacterium dehalogenans ATCC 51507	Desde_0044	AFL98539	none	690
Desulfitobacterium dehalogenans ATCC 51507	Desde_4096	AFM02357	none	403
Dethiobacter alkaliphilus AHT 1	DealDRAFT_0257	EEG78983	none	407
Acidobacterium capsulatum ATCC51196	ACP_0433	ACO34110	none	330
Gamma proteobacterium sp. strain NOR51-B	NOR51B_192	EED34255	none	328
Gamma proteobacterium sp. strain NOR51-B	NOR51B_2467	EED36515	none	332
Gamma proteobacterium sp. strain NOR51-B	NOR51B_2552	EED36600	none	391
Gamma proteobacterium sp. strain NOR51-B	NOR51B_2836	EED36883	none	449
Rhodobacterales uncultured bacterium HTCC2083	RB2083_1520	EDZ42005	none	1068
Rhodobacterales uncultured bacterium HTCC2083	RB2083_1633	EDZ42118	none	390
Rhodobacterales uncultured bacterium Y4I	RBY4I_2190	EDZ46972	none	390
Rhodobacterales bacterium Y4I	RBY4I_3858	EDZ48635	none	1047
Pseudovibrio sp. strain JE062	PJE062_4085	EEA92265	none	1074
Ruegeria sp. strain R11	RR11_825	EEB70068	none	1102
Roseobacter sp. strain GAI101	RGAI101_532	EEB83384	none	1067
Rhodobacteraceae uncultured bacterium KLH11	RKLH11_4169	EEE35492	none	388
Gamma proteobacterium sp. strain NOR5-3	NOR53_1610	EED32850	none	1074
Clostridium difficile QCD-66c26	CdifQC_020100009594	ABFD00000000	none	354
Thalassiobium sp. strain R2A62	TR2A62_2026	EET46325	none	1044
Clostridium difficile ATCC43255	CdifA_020200010217	ABKJ00000000	none	354
Clostridium difficile QCD-23m63	CdifQCD-2_020200009469	ABKL00000000	none	354
Silicibacter sp. strain TrichCH4B	SCH4B_0549	EEW60663	none	1070
Silicibacter lacuscaerulensis ITI-1157	SL1157_0409	EEX08396	none	1070
Ferroglobus placidus DSM 10642	Ferp_2321	ADC66443	none	406
Ahrensia sp. strain R2A130	R2A130_1084	EFL88602	none	1068
Thermotogales uncultured bacterium mesG1.Ag.4.2	ThebaDRAFT_2271	.	none	312
Roseibium sp. strain TrichSKD4	TRICHSKD4_0494	EFO34197	none	1040
Roseobacter litoralis Och 149	RLO149_c003740	AEI92404	none	1051
Ruegeria sp. strain TW15	RTW15_010100002114	.	none	388
Ruegeria sp. strain TW15	RTW15_9921		none	1070
Thermovirga lienii DSM17291	Tlie_0302	AER66041	none	355
Clostridium difficile 050-P50-2011	HMPREF1123_02575	EHJ27256	none	345
Flavonifractor plautii ATCC 29863	HMPREF0372_00086	EHM55262	none	400
uncultured marine bacterium 105	MBMO_EBAC750-01A01.31	AAR37457	none	396
Ruegeria pomeroyi DSS-3	SPO0143	AAV93471	none	328
Ruegeria pomeroyi DSS-3	SPO0589	AAV93904	none	1090
Ruegeria pomeroyi DSS-3	SPO1738	AAV95018	none	394
Sulfitobacter sp. strain EE-36	EE36_11189	EAP83333	none	1047
Sulfitobacter sp. strain NAS-14.1	NAS141_16409	EAP79417	none	1047
Roseobacter sp. strain MED193	MED193_08608	EAQ45701	none	395
Jannaschia sp. strain CCS1	Jann_1968	ABD54885	none	399
Jannaschia sp. strain CCS1	Jann_3974	ABD56891	none	1069
Ruegeria sp. strain TM1040	TM1040_3578	ABF62547	none	1070
Roseobacter denitrificans OCh 114	RD1_4098	ABG33540	none	1070

A.11 Table S5. Sequences used to build non-RDase HMMs. A. Polyferredoxin sequences served as a negative model in building a non-RDase iron-sulfur protein HMM and represented the only enzymologically-characterized non-RDases retrieved by the RDase HMM (bit scores < 110). **B.** A second negative model was built from characterized iron-sulfur protein sequences from genomes or operons of non-organohalide-respiring model organisms.

Organism	annotation	Accession (locus)	References
	A		
Methanothermobacter marburgensis str. Marburg	Polyferredoxin protein mvhB	P60232 (MTBMA_c15160)	(Hedderich et al., 1992)
Methanothermus fervidus DSM 2088	Polyferredoxin protein mvhB	Q49180	(Steigerwald et al., 1990)
Methanothermobacter thermautotrophicus str. Delta H	Polyferredoxin protein mvhB	Q50784 (MTH_1133)	None
Methanococcus voltae ATCC 33273	Polyferredoxin protein vhuB	Q00388	None
Methanocaldococcus jannaschii DSM 2661	Polyferredoxin protein fwdF	Q58566 (MJ1166)	None
Methanocaldococcus jannaschii DSM 2661	Uncharacterized polyferredoxin-like protein	P81293 (MJ0514.2)	None
	B		
Escherichia coli K-12	Putative electron transport protein yjeS	P39288 (b4166)	(Blattner et al., 1997)
Geobacillus stearothermophilus subsp. *nondiastaticus*	Uncharacterized protein in gldA 3'region	P32815	(Mallinder et al., 192)
Bacillus subtilis subsp. *subtilis* str. 168	Putative electron transport protein yhbA	P97030 (BSU08910)	(Kunst et al., 1997)

A.12 Table S6: Source and range of clone accessions for non-genomic RDases retrieved by the RDase HMM. Class 1 sequences retrieved by RDase HMM from from enrichment culture or environmental isolate clone RDases exhibited strong similarity with the RDase HMM (avg. bit scores > 300). Sequence lengths denote number of amino acid residues on amplicon RDase ORF while length % HMM match indicates proportions of retrieved sequences covered by RDase HMM.

Source/Study	Clone Accessions	Number of sequences	Average bit score	Sequence lengths (aa)	Length % HMM match
Dehalococcoides enrichment culture clones KS/RC (Ritalahti and Löffler, 2004)	ABY28306 - ABY28335	29 class 1	366	460 - 504	97 - 100
Dehalococcoides culture clones FL2, CBDB1, BAV1 (Hölscher, et al, 2004)	AAR24296 - AAR24325, AAT48548 - AAT48560, AAZ91690 - AAZ91692	32 class 1	347	449 - 507	97 - 100
TCE-degrading enrichment culture clones (Krajmalnik-Brown, et al., 2007)	AAN85588 - AAN85594 ABB89703 - ABB89715	8 class 1	621	526 - 554	all 100
TCE-degrading groundwater samples (Lee et al., 2008)	ABV44386 - ABV44392	4 class 1	447	465 - 509	97 - 100
Dehalococcoides strain MB culture (Chow et al., 2010)	ACF24861- ACF24863	2 class 1	442	489 - 496	98 - 100
KB1 consortium (Duhamel et al., 2004)	ABA64523 - ABA64549	14 class 1	411	459 - 503	97 - 100
Pentachlorobenzene-dechlorinating sediment (Santoh et al., 2006)	BAF34634 - BAF34633	4 class 1	348	457 - 492	97 - 100
Groundwater (Nakamura et al., 2006)	BAF34974 - BAF34982	8 class 1	623	540 - 544	99 - 100
TUT2264 consortium (Futamata et al., 2009)	BAG72156 - BAG72169	11 class 1	322	466 - 504	97 - 99
TCE-contaminated groundwater (Nishimura et al., 2008)	BAF45397 - BAF45413	11 class 1	379	445 - 519	97 - 100
pristine seafloor sediment, some bromophenol-degrading (Futagami et al., 2009)	BAI47842 - BAI47804	32 class 1	313	442 - 550	96 - 100
1,2-DCA-degrading-consortium (Marzorati et al., 2010)	CAR57926 - CAR57937	11 class 1	559	all 551	100
1,2-DCA-degrading-consortium (Marzorati, et al., 2007)	CAJ75435	1 class 1	562	551	100

A.13 Table S7. Class 1 clades. Clustering of class 1 sequences into evolutionary lineages in the 459-sequence MSA, 395 out of 406 total.

Clade	Characterized RDases in clade	Genomes/isolates represented	Bootstrap support at ancestral node	Sequence count	Global amino acid sequence pairwise identities	
					range	> 40% to known RDase
a	PceA, TceA, VcrA, BvcA, and CbrA	*Dehalococcoides mccartyi* Strains 195, CBDB1, BAV1, VS, GT, FL2, and MB. *Dehalogenimonas lykanthroporepellens* strain BL-DC-9, Candidatus Dehalobium chlorocoercia strain DF-1	45.0%	309	22% to 100%	75
b	3 PceAs and CprA5	*Desulfitobacterium hafniense* strains Y51, DCB-2, and PCP-1; *Desulfitobacterium* sp. strain PCE-S. *Dehalobacter* sp. strains PER, WL, and *Dehalobacter* metagenome CF-50. *Geobacter lovleyi* strain SZ	57.0%	24	33% to 99%	20
c	4 CprAs	*Desulfitobacterium chlororespirans, Desulfitobacterium dehalogenans, Desulfitobacterium hafninese* DCB-2, and *Desulfitobacterium* sp. strains KBC1, Viet1, and PCE1, and *Dehalobacter* metagenome CF-50	94.5%	19	52% to 99%	15
d	None	*Dehalococcoides mccartyi* strains 195, CBDB1, BAV1, VS, and GT	100.0%	7	55% to 100%	0
e	None	*Dehalococcoides mccartyi* strains 195, CBDB1, BAV1, and VS	98.5%	5	21% to 99%	0
f	None	*Shewanella sediminis* HAW-EB3, *Vibrio* sp. strain RC586, and *Vibrio scopthalmii* LMG19158	73%	6	38% to 81%	0
g	None	*Desulfitobacterium hafniense* DCB-2, *Dehalobacter* metagenome CF-50, *Shewanella sediminis* HAW-EB3, *Photobacterium profundum* 3TCK, *Vibrio ichthyoenteri* ATCC700023, *Vibrio scopthalmii* LMG19158, and *Vibrio* sp. strain N418.	43.5%	7	29% to 97%	0
h	None	*Anaeromyxobacter dehalogenans* strains 2CP-C, 2CP-1, and K.	100%	3	86% to 97%	0
i	None	*Dehalobacter* metagenome CF-50	97%	3	59% to 63%	0
j	PceA	*Sulfurospirillum multivorans*	100%	2	91%	1
k	None	*Anaeromyxobacter dehalogenans* strains 2CP-C, 2CP-1, and K.	100%	3	89% to 99%	None
single	PceA (PrdA)	*Desulfitobacterium* sp. strain KBC1	n/a	1	-	-
single	None	*Dehalobacter* metagenome CF-50				
single	None	*Deltaproteobacterium* sp. strain NaphS2	n/a	1	-	-
single	None	*Dehalobacter* metagenome CF-50	n/a	1	-	-

A.14 Table S8. Class 2 clades. Clustering of class 2 sequences into evolutionary lineages in the 459-sequence MSA - 55 out of 59 total.

Clade	Characterized RDases in clade	Genomes/isolates represented	Bootstrap support at ancestral shared node	Sequence count	Global amino acid sequence pairwise identities	
					range	> 40% to known RDase
a	None	*Ahrensia* sp. strain R2A130, *Jannaschia* sp. strain CCS2, *Oceanibulbus indolifex* HEL-45, *Phaeobacter gallaeciensis*, *Pseudovibrio* sp. strain JE062, unclassified *Rhodobacterales* spp. strains HTCC2083 and Y41, *Roseobacter denitrificans* and *litoralis*, *Roseobacter* spp. strains CCS2, GAI101, and SK209-2-6, *Ruegeria pomeroyi* DSS-3, *Ruegeria* spp. strains R11 and TM1040, *Silicibacter lacuscaerulensis* and *Silicibacter* sp. strain TrichCH4B, *Sulfitobacter* spp. strains EE-36 and NAS-14-1, *Thalassiobium* strain R2A62, unclassified *α-Proteobacteria* spp. strains HTCC2255 and HTCC2150; *γ-Proteobacterium* sp. strain NOR5-3, and *Rosiebium* sp. strain TrichSKD4,	100%	25	45% to 100%	None
b	None	unclassified *Rhodobacterales* spp. strains KLH11, HTCC2083, and Y41; *γ-Proteobacterium* sp. strain NOR51-3, *Ruegeria pomeroyi* DSS-3, *Jannaschia* sp. strain CCS2, *Roseobacter* sp. strain SK209-2-6, *Rosiebium* sp. strain TrichSKD4, and uncultured strains 105 and HF4000 010I05	82%	10	36% to 80%	None
c	None	*Acidobacterium capsulatum* ATCC51196, *Cyanobium* sp. strain PCC7001, *γ-Proteobacterium* spp. strains NOR51-3 and HTCC2080, *Ruegeria pomeroyi* DSS-3	72.5%	6	24% to 67%	None
d	None	*Clostridium bartlettii* DSM16795, *Clostridium difficile* strains 630, QCD-66c26, QCD-23m63, and ATCC43255, and *Thermovirga leinii* DSM17291	49.5%	7	70% to 100%	None
e	None	*Heliobacterium modesticaldum* Ice1 and *Dethiobacter alkaliphilus* AHT 1	69%	2	55%	None
f	None	Uncultured marine microorganism HF4000_010I05	100%	2	87%	None
single	None	*Desulfitobacterium hafniense* strain DCB-2	-	1		None
single	None	*Ferroglobus placidus* DSM 10642	-	1	-	-
single	None	Uncultured marine microorganism HF4000	-	1	-	-
single	None	*Flavonifractor plautii* ATCC 29863	-	1	-	-

A.15 Chapter 3, Supplemental References

Adrian, L., Rahenführer, J., Gobom, J., and Hölscher, T. (2007) Identification of a Chlorobenzene reductive dehalogenase in *Dehalococcoides* sp. Strain CBDB1. *Applied and Environmental Microbiology* **73**: 7717-7724.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M. et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1462.

Chow, W.L., Cheng, D., Wang, S. and He, J. (2010) Identification and transcriptional analysis of *trans*-DCE-producing reductive dehalogenases in *Dehalococcoides* species. *ISME Journal* **4**: 1020-1030.

Duhamel, M., Mo, K., and Edwards, E.A. (2004) Characterization of a highly enriched *Dehalococcoides*-containing culture that grows on vinyl chloride and trichloroethene. *Applied and Environmental Microbiology* **70**: 5538-5545.

Futagami, T., Morono, Y., Terada, T., Kaksonen, A.H., and Inagaki, F. (2009) Dehalogenation activities and distribution of reductive dehalogenase homologous gene in marine subsurface sediments. *Applied and Environmental Microbiology* **75**: 6905-6909.

Futamata, H., Kaiya, S., Sugawara, M., and Hirashi, A. (2009) Phylogenetic and transcriptional analysis of a tetrachloroethene-dechlorinating "*Dehalococcoides*" enrichment culture TUT2264 and its reductive-dehalogenase genes. *Microbes and Environments* **24**: 330-337.

Hedderich, R., Albracht, S. P. J., Linder, D., Koch, J., and Thauer, R. K. (1992) Isolation and characterization of polyferredoxin from *Methanobacterium thermoautotrophicum*. The mvhB gene product of the methylviologen-reducing hydrogenase operon. *FEBS Letters* **298**:65-68.

Hölscher, T., Krajmalnik-Brown, R., Ritalahti, K.M., Wintzingerode, F.v., Görisch, H., Löffler, F.E., and Adrian, L. (2004) Multiple nonidentical reductive-dehalogenase-homologous genes are common in *Dehalococcoides*. *Applied and Environmental Microbiology* **70**: 5290-5297.

Kim, S.-H., Harzman, C., Davis, J.K., Hutcheson, R., Broderick, J.B., Marsh, T.L., Tiedje, J.M. (2012) Genome sequence of *Desulfitobacterium hafniense* DCB-2, a Gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiology* **12:** 21.

Krajmalnik-Brown, R., Sung, Y., Ritalahti, K.M., Saunders, F.M., and Löffler, F.E. (2007) Environmental distribution of the trichloroethene reductive dehalogenase gene (*tceA*) suggests lateral gene transfer among *Dehalococcoides*. *FEMS Microbiology Ecology* **59**: 206-214.

Krajmalnik-Brown, R., Hölsher, T., Thomson, I.N., Saunders, F.M., Ritalahti, K.M., and Löffler, F.E. (2004) Genetic identification of a putative vinyl chloride reductase in *Dehalococcoides* sp. Strain BAV1. *Applied and Environmental Microbiology* **70**: 6347-6351.

Krasotkina, J., Walters, T., Maruya, K.A., and Ragsdale, S.W. (2001) Characterization of the B_{12}- and iron-sulfur-containing reductive dehalogenase from *Desulfitobacterium chlororespirans*. *The Journal of Biological Chemistry* **276**: 40991-40997.

Kube, M., Beck, A., Zinder, S.H., Kuhl, H., Reinhardt, R., and Adrian, L. (2005) Genome sequence of the chlorinated compound-respiring bacterium *Dehalococcoides* species strain CBDB1. *Nature Biotechnology* **23**: 1269-1273.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **890**:249-256.

Lee, P.K., Macbeth, T.W., Sorenson, K.S. Jr., Deeb,R.A. and Alvarez-Cohen,L. (2008) Quantifying genes and transcripts to assess the in situ physiology of '*Dehalococcoides*' spp. in a trichloroethene-contaminated groundwater site. *Applied and Environmental Microbiology* **74**: 2728-2739.

Magnuson, J.K., Romine, M.F., Burris, D.R., and Kingsley, M.T. (2000) Trichloroethene reductive dehalogenase from *Dehalococcoides ethenogenes*: Sequence of *tceA* and substrate range characterization. *Applied and Environmental Microbiology* **66**: 5141-5147.

Magnuson, J.K., Stern, R.V., Gosset, J.M., Zinder, S.H., and Burris, D.R. (1998) Reductive dechlorination of tetrachloroethene to ethane by a two-component enzyme pathway. *Applied and Environmental Microbiology* **64**: 1270-1275.

Maillard, J., Regeard, C., and Holliger, C. (2005) Isolation and characterization of Tn-Dha1, a transposon containing the tetrachloroethene reductive dehalogenase of *Desulfitobacterium hafniense* strain TCE1. *Environmental Microbiology* **7**: 107-117.

Maillard, J., Schumacher, W., Vazquez, F., Regeard, C., Hagen, W.R., and Holliger, C. (2003) Characterization of the corrinoid iron-sulfur protein tetrachloroethene reductive dehalogenase of *Dehalobacter restrictus*. *Applied and Environmental Microbiology* **69**: 4628-4638.

Mallinder, P. R., Pritchard, A., and Moir, A. (1992) Cloning and characterization of a gene from *Bacillus stearothermophilus* var. *non-diastaticus* encoding a glycerol dehydrogenase. *Gene* **110**:9-16.

Marzorati, M., Balloi, A., de Ferra, F., Corallo, L., Carpani, G., Wittebolle, L., Verstraete, W. and Daffonchio, D. (2010) Bacterial diversity and reductive dehalogenase redundancy in a 1,2-dichloroethane-degrading bacterial consortium enriched from a contaminated aquifer. *Microbial Cell Factories* **12**.

Marzorati, M., Ferra, F.d., Raemdonck, H.V., Borin, S., Allifranchini, E., Carpani, G. et al. (2007) A novel reductive dehalogenase, identified in a contaminated groundwater enrichment culture and in *Desulfitobacterium dichloroeliminans* strain DCA1, is linked to dehalogenation of 1,2-Dichloroethane. *Applied and Environmental Microbiology* **73**: 2990-2999.

McMurdie, P.J., Behrens, S., Müller, J.A., Göke, J., Ritalahti, K.M., Wagner, R. et al. (2009) Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. *PLoS Genetics* **5**: e1000714.

Miller, E., Wohlfarth, G., and Diekert, G. (1998) Purification and characterization of the tetrachloroethene reductive dehalogenase of strain PCE-S. *Archives of Microbiology* **169**: 497-402.

Mohn, W. W. and Kennedy, K. J. (1992) Reductive dehalogenation of chlorophenols by *Desulfomonile tiedjei* DCB-1. *Applied and Environmental Microbiology* **58**:1367-1370.

Müller, J.A., Rosner, B.M., Abendroth, G.v., Meshulam-Simon, G., McCarty, P.L., and Spormann, A.M. (2004) Molecular identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. Strain VS and its environmental distribution. *Applied and Environmental Microbiology* **70**: 4880-4888.

Nakamura, K., Mizumoto, M., Ueno, T. and Ishida, H. (2006) Cloning and analysis of trichloroethene reductive dehalogenase gene and its detection by quantitative real-time PCR. *Environmental Engineering Research* **43**: 119-125.

Neumann, A., Wohlfarth, G., and Diekert, G. (1996) Purification and characterization of tetrachloroethene reductive dehalogenase from *Dehalospirillum multivorans*. *The Journal of Biological Chemistry* **271**: 16515-16519.

Nishimura, M., Ebisawa, M., Sakihara, S., Kobayashi, A., Nakama, T., Okochi, M., and Yohda, M. (2008) Detection and identification of *Dehalococcoides* species responsible for *in situ* dechlorination of trichloroethene to ethene enhanced by hydrogen-releasing compounds. *Biotechnology and Applied Biochemistry* **51**: 1-7.

Nonaka, H., Keresztes, G., Shinoda, Y., Ikenaga, Y., Abe, M., Inatomi, K., Furukawa, K., Inui, M., and Yukawa, H. (2006) Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195. *Journal of Bacteriology* **188:** 2262-2274.

Ritalahti, K.M., and Löffler, F.E. (2004) Populations implicated in anaerobic reductive dechlorination of 1,2-dichloropropane in highly enriched bacterial communities. *Applied and Environmental Microbiology* **70**: 4088-4095.

Sanford, R.A., Cole, J.R., and Tiedje, J.M. (2002) Characterization and description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an aryl-halorespiring facultative anaerobic myxobacterium. *Applied and Environmental Microbiology* **68**: 893-900.

Santoh, K., Kouzuma, A., Ishizeki, R., Iwata, K., Shimura, M., Hayakawa, T., Hoaki, T., Nojiri, H., Omori, T., Yamane, H. and Habe, H. (2006) Detection of a bacterial group within the phylum *Chloroflexi* and reductive-dehalogenase-homologous genes in pentachlorobenzene-dechlorinating estuarine sediment from the Arakawa River, Japan. *Microbes Enivronment* **21**: 154-162.

Seshadri, R., Adrian, L., Fouts, D.E., Eisen, J.A., Phillippy, A.M., Methe, B.A. et al. (2005) Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* **307**: 105-108.

Steigerwald, V. J., Beckler, G. S., and Reeve, J. N. (1990) Conservation of hydrogenase and polyferredoxin structures in the hyperthermophilic archaebacterium *Methanothermus fervidus*. *Journal of Bacteriology* **172**:4715-4718.

Sung, Y., Ritalahti, K. M., Apkarian, R. P., and Löffler, F.E. (2006) Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring *Dehalococcoides* isolate. *Applied and Environmental Microbiology* **72**:1980-1987.

Suyama, A., Yamashita, M., Yoshino, S., and Furukawa, K. (2002) Molecular characterization of the PceA reductive dehalogenase of *Desulfitobacterium* sp. Strain Y51. *Journal of Bacteriology* **184**: 3419-3425.

Thibodeau, J., Gauthier, A., Duguay, M., Villemur, R., Lépine, F., Juteau, P., and Beaudet, R. (2004) Purification, cloning, and sequencing of a 3,5-dichlorophenol reductive dehalogenase from *Desulfitobacterium frappieri* PCP-1. *Applied and Environmental Microbiology* **70**: 4532-4537.

Thomas, S.H., Wagner, R.D., Arakaki, A.K., Skolnick, J., Kirby, J.R., Shimkets, L.J., Sanford, R.A., Löffler, F.E. (2008) The mosaic genome of *Anaeromyxobacter dehalogenans* strain 2CP-C suggests an aerobic common ancestor to the delta-proteobacteria. *PLoS One* **3:** e2103.

Tsukagoshi, N., Ezaki, S., Uenaka, T., Suzuki, N., and Kurane, R. (2005) Isolation and transcriptional analysis of novel tetrachloroethene reductive dehalogenase gene from *Desulfitobacterium* sp. strain KBC1. *Applied Microbiology and Biotechnology.* **69**: 543-553.

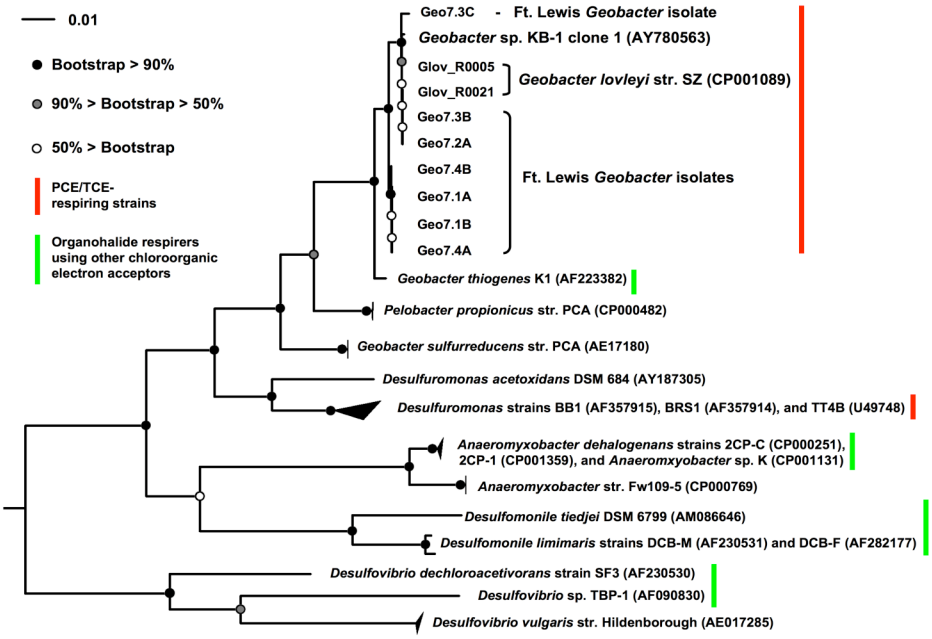
Utkin, I., Woese, C., and Wiegel, J. (1994) Isolation and characterization of *Desulfitobacterium dehalogenans* gen. nov., sp. nov., an anaerobic bacterium which reductively dechlorinates chlorophenolic compounds. *International Journal of Systematic Bacteriology* **44**: 612-619.

van den Pas, B.A.v.d., Smidt, H., Hagen, W.R., Oost, J.v.d., Schraa, G., Stams, A.J.M., and Vos, W.M.d. (1999) Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in *Desulfitobacterium dehalogenans*. *The Journal of Biological Chemistry* **274**.

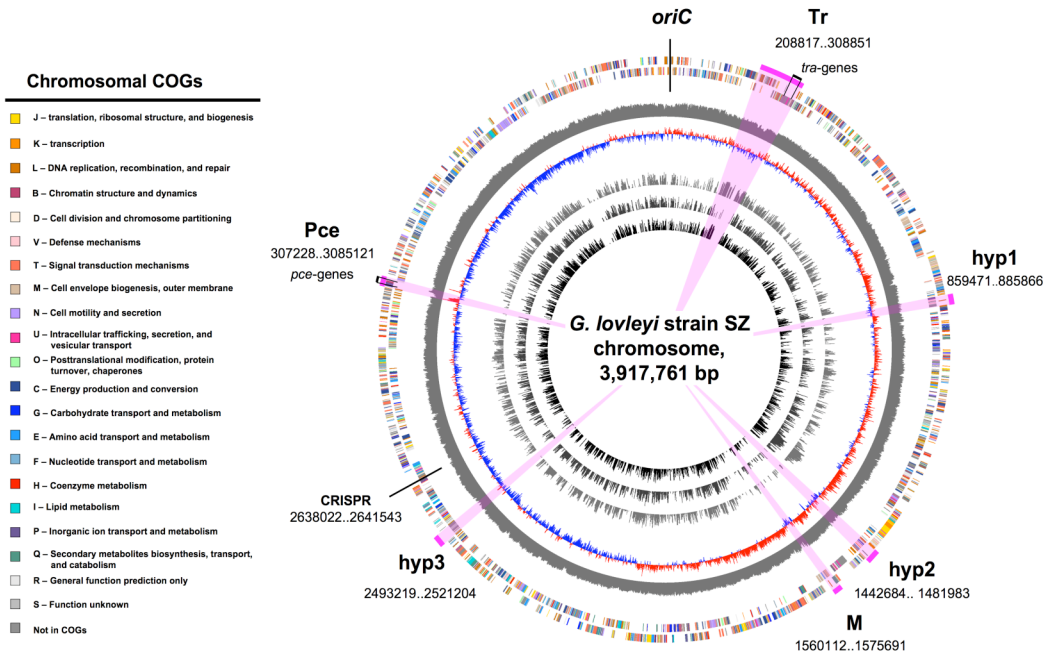
Wagner, A., Adrian, L., Kleinsteuber, S., Andreesen, J. R., Lechner, U. (2009) Transcription analysis of genes encoding homologues of reductive dehalogenases in "*Dehalococcoides*" sp. Strain CBDB1 by using terminal restriction fragment length polymorphism and quantitative PCR. *Applied and Environmental Microbiology* **75**:1876-1844.

Wagner, D.D., Hug, L.A., Hatt, J.K., Spitzmiller, M.R., Padilla-Crespo, E., Ritalahti, K.M. et al. (2012) Genomic determinants of organohalide-respiration in *Geobacter lovleyi*, an unusual member of the *Geobacteraceae. BMC Genomics* **13**: 200.

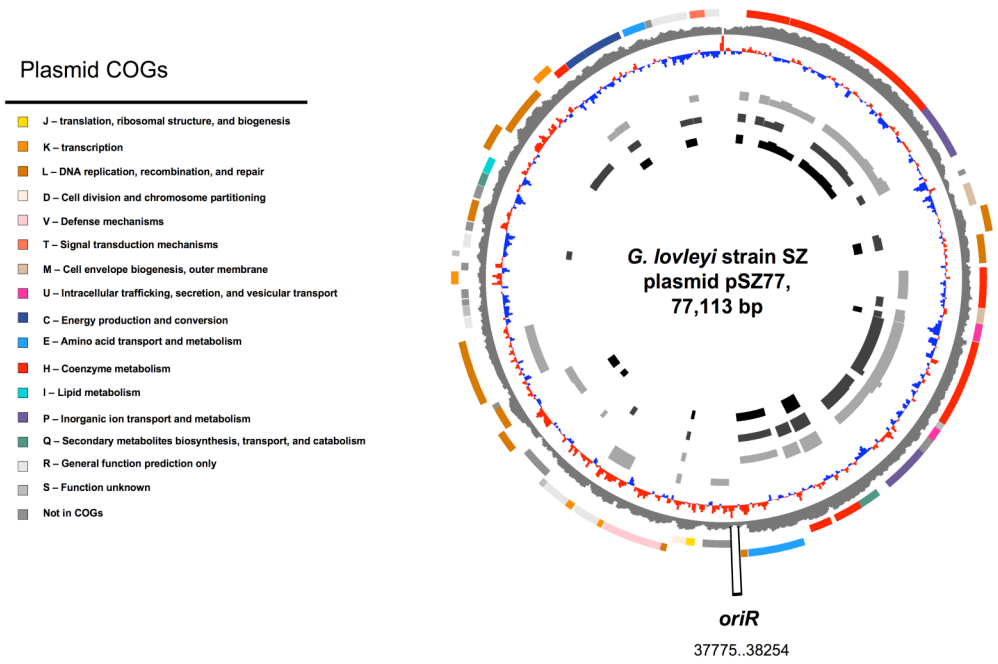
Yan, J., Rash, B. A., Rainey, F. A., Moe, W. M. (2009) Isolation of novel bacteria within the *Chloroflexi* capable of reductive dechlorination of 1,2,3-trichloropropane. *Environmental Microbiology* 11: 833-843.



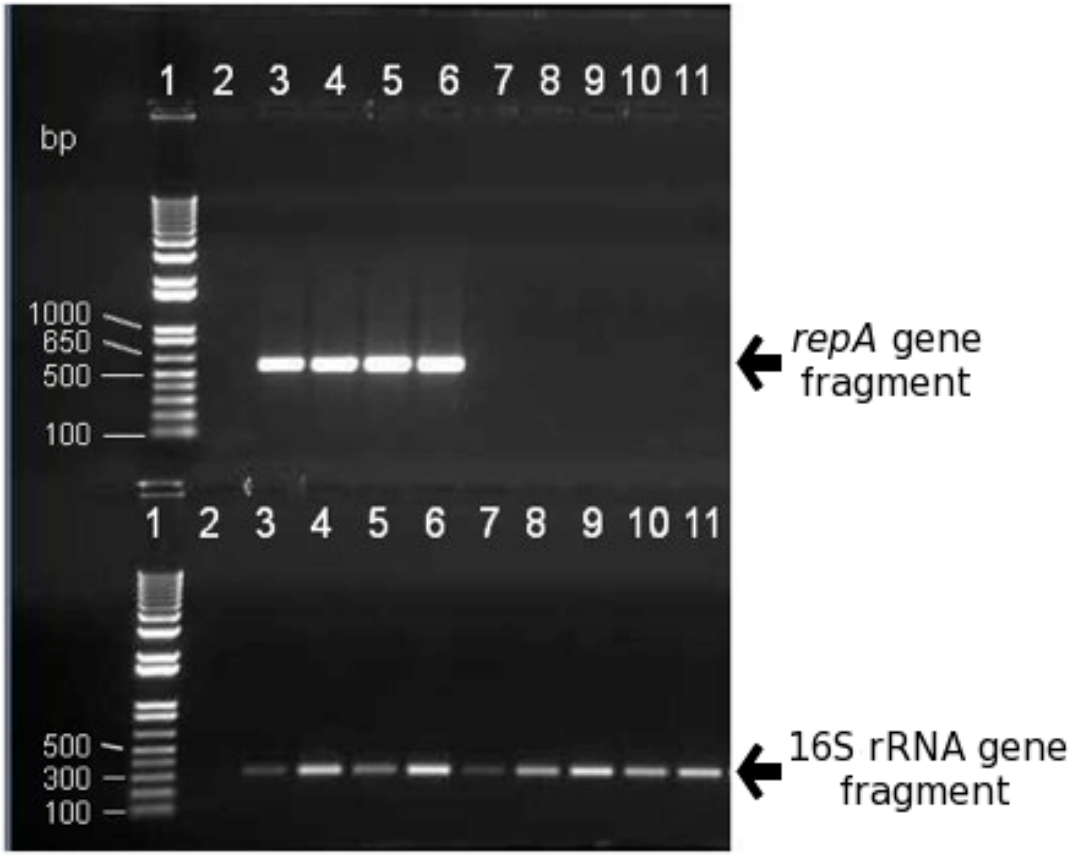
B.1 Figure S1. Unrooted 16S rRNA gene tree showing organohalide-respiring *δ-Proteobacteria* from isolates or mixed cultures. Selected related non-dechlorinating organisms (non-highlighted branches) are added to show divergence within genera.



B.2 Figure S2. Circular genome map of the *G. lovleyi* strain SZ chromosome (CP001089). From outside to center: COG categories of genes on forward strand, COG categories of genes on reverse strand, percent G+C content, GC skew, and percent Blastx identity of strain SZ ORFs to ORFs on the chromosomes of *Pelobacter propionicus* (CP000482), *G. uraniireducens*, and *G. sulfurreducens*. Genomic island regions, six in all, are indicated by pink arcs and are characterized by genes lacking blast matches in a majority of other *Geobacter*/*Pelobacter* spp. genomes (e.g. the three innermost circles), suggesting horizontal gene transfer. The genomic island Pce harbors the *pce*-gene cluster, encoding the PCE reductive dehalogenases, and exhibits nucleotide sequence deviations in G+C% and GC-skew relative to the entire strain SZ chromosome. Chromosome map created using GenomeViz (Ghai et al., 2004); View at 200%.



B.3 Figure S3. Circular genome map of the *G. lovleyi* strain SZ plasmid pSZ77 (CP001090). From outside to center: COG categories of genes on forward strand, COG categories of genes on reverse strand, percent G+C content, GC skew, percent blastx identity of SZ ORFs to plasmids of *Pelobacter propionicus* (CP000483 and CP000484), and percent Blastx identity to the chromosomes of *Pelobacter propionicus* (CP000482), *Geobacter uraniireducens*, and *G. sulfurreducens*. Plasmid map created using GenomeViz (Ghai et al., 2004).



B.4 Figure S4. Plasmid pSZ77 *repA* gene and *Geobacteraceae* spp. 16S rRNA genes in pure and mixed cultures containing *Geobacter* strains with specific PCR primers. Top row: *repA* gene-targeted PCR (565 bp amplicon) and bottom row *Geobacteraceae* 16S rRNA gene-targeted PCR (312 bp amplicon). The arrows indicate the expected PCR amplicons. Lane 1: 1 kb Plus DNA ladder (Invitrogen), Lane 2: no template (negative control), Lane 3: *G. lovleyi* strain SZ (positive control), Lanes 4 and 5: consortium BDI, Lane 6: consortium KB-1, Lane 7: Ft. Lewis isolate 7.1, Lane 8: Ft. Lewis isolate 7.2, Lane 9: Ft. Lewis isolate 7.3, Lane 10: Ft. Lewis isolate 7.4, Lane 11: *G. thiogenes*.

B.5 Table S1. Pce genomic island CAI. Codon usage of *pce*-genes and adjacent genes encoded on the Pce genomic island in comparison to all strain SZ chromosomal genes using the codon adaptation index (CAI). Normalized CAI < 1.00 (red font) indicates a possible laterally acquired gene and is scored below genomic expected CAI at a 5% level of significance (genomes.urv.es/CAIcal/E-CAI).

SZ chromosome locus	Function	Length (bp)	Normalized CAI*
Glov_2865	IstB domain ATP-binding protein	744	1.03
Glov_2866	Integrase catalytic region	1512	1.11
Glov_2868	PceT - Peptidylprolyl isomerase	936	0.92
Glov_2869	PceC – FMN-binding domain protein	1149	0.85
Glov_2870	PceA – reductive dehalogenase, catalytic subunit	1545	0.87
Glov_2871	PceB – reductive dehalogenase membrane anchor subunit	330	0.85
Glov_2872	PceA – reductive dehalogenase, catalytic subunit	1545	0.87
Glov_2873	PceB – reductive dehalogenase membrane anchor subunit	330	0.85
Glov_2874	Aldehyde dehyrogenase	1416	0.95
Glov_2875	Transposase IS204/IS1001/IS1096/IS1165	1176	0.96

* Normalized to codon usage over the entire strain SZ chromosome

B.6 Table S2. Tra genomic island CAI. Codon usage of predicted F-factor conjugative pilus *tra*-genes in comparison to all strain SZ chromosomal genes using the codon adaptation index (CAI). A majority of genes in the *tra*-pilus cluster have normalized CAI < 1.00 (red font), but are interspersed with genes having normalized CAI > 1.00, indicating sufficient residence time in strain SZ for the cluster to partially ameliorate to chromosomal codon usage.

SZ chromosome locus	Function with predicted PFAM domains	Length (bp)	Normalized CAI
Glov_0304	Conserved hypothetical protein	324	0.85
Glov_0305	TraG – DNA transfer and F pilus assembly protein - pfam07916	3618	0.99
Glov_0306	Conserved hypothetical protein	741	1.01
Glov_0307	Conserved hypothetical protein	252	0.96
Glov_0308	TraF – type IV secretory protease – pfam10502	495	1.09
Glov_0309	TrbI – Conserved hypothetical protein – pfam09677	393	0.99
Glov_0310	TrbC – pilin assembly protein – pfam09673	963	1.02
Glov_0311	TraU – TraU family protein – pfam06834	1023	1.01
Glov_0312	Hypothetical protein, no NCBI database hits	471	0.89
Glov_0313	TraW - Conserved hypothetical cytosolic protein (TraW – TIGR02743)	657	0.98
Glov_0314	TraN - Hypothetical protein (TraN_Ftype – TIGR02750)	3375	1.00
Glov_0315	Conserved hypothetical protein	2004	1.07
Glov_0316	TraC – sex pilus assembly protein – pfam11130	2448	1.07
Glov_0317	TraV – conserved hypothetical protein – pfam09676	495	1.05
Glov_0318	TrbI – TraB pilus assembly family protein – (TrbI) pfam03743	1269	0.95
Glov_0319	TraK – conserved hypothetical protein – pfam06586	948	0.98
Glov_0320	TraE – conserved hypothetical sex pilus assembly and synthesis protein - pfam05309	579	1.08
Glov_0321	TraL – conserved hypothetical protein – pfam07178	267	0.97
Glov_0322	TrbC – conserved hypothetical protein – pfam04956	300	0.97

* Normalized to codon usage over the entire strain SZ chromosome

B.7 Table S3. Inferred *c*-type cytochrome genes on the *G. lovleyi* strain SZ chromosome.

Locus	Gene symbol	CxxCH motifs	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.*
Glov_0202		2	YP_383296	*Geobacter metallireducens* GS-15	55	183/255
Glov_0209		3	YP_383303	*Geobacter metallireducens* GS-15	74	72/93
Glov_0211	*nrfA*	5	ZP_05311341	*Geobacter* sp. M18	72	375/445
Glov_0642		2	YP_001230785	*Geobacter uraniireducens* Rf4	60	435/599
Glov_0860		8	YP_901261	*Pelobacter propionicus* DSM 2379	71	400/470
Glov_0887		4	YP_001365373	*Shewanella baltica* OS185	40	64/106
Glov_0942		6	YP_002135959	*Anaeromyxobacter* sp. K	58	58/84
Glov_0946		4	ZP_01313446	*Desulfuromonas acetoxidans* DSM684	49	73/114
Glov_1004		6	YP_002135959	*Anaeromyxobacter* sp. K	56	61/85
Glov_1042	*nrfA*	5	YP_899549	*Pelobacter propionicus* DSM 2379	70	384/462
Glov_1043	*nrfH*	4	YP_383265	*Geobacter metallireducens* GS-15	60	115/147
Glov_1044		8	ZP_05311062	*Geobacter* sp. M18	60	298/406
Glov_1051		3	YP_001232971	*Geobacter uraniireducens* Rf4	71	52/63
Glov_1059		2	YP_004196830	*Geobacter* sp. M18	54	64/92
Glov_1061		7	ZP_05311265	*Geobacter* sp. M18	71	447/544
Glov_1150		3	YP_001229258	*Geobacter uraniireducens* Rf4	58	70/101
Glov_1172		5	YP_383502	*Geobacter metallireducens* GS-15	59	233/316
Glov_1177		8	YP_003020606	*Geobacter* sp. M21	69	384/477
Glov_1194		4	YP_002536809	*Geobacter* sp. FRC-32	46	148/251
Glov_1198	*mtrF*	10	YP_002140812	*Geobacter bemidjiensis* Bem	57	297/413
Glov_1199		12	ZP_05310895	*Geobacter* sp. M18	60	456/593
Glov_1201		5	YP_001229236	*Geobacter uraniireducens* Rf4	63	221/302
Glov_1229		1	NP_954374	*Geobacter sulfurreducens* PCA	51	55/81
Glov_1291		6	YP_003502960	*Denitrovibrio acetiphilus* DSM 12809	63	108/139
Glov_1315		12	YP_003021711	*Geobacter* sp. M21	57	237/333
Glov_1467		9	ZP_05310213	*Geobacter* sp. M18	47	382/604
Glov_1468		9	ZP_05310213	*Geobacter* sp. M18	47	404/642
Glov_1483		1	NP_952450	*Geobacter sulfurreducens* PCA	70	473/561
Glov_1703		12	NP_951650	*Geobacter sulfurreducens* PCA	50	200/324
Glov_1706		5	YP_001232378	*Geobacter uraniireducens* Rf4	52	119/175
Glov_1710		5	YP_900554	*Pelobacter propionicus* DSM 2379	61	379/490
Glov_1719		8	YP_901261	*Pelobacter propionicus* DSM 2379	72	390/471

B.7 Table S3. (continued)

Locus	Gene symbol	CxxCH motifs	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.*
Glov_1762		3	ZP_01312357	*Desulfuromonas acetoxidans* DSM684	49	372/571
Glov_1764		8	YP_003020606	*Geobacter* sp. M21	66	376/480
Glov_1803		6	ZP_05313793	*Geobacter* sp. M18	70	318/393
Glov_1903		1	YP_001230676	*Geobacter uraniireducens* Rf4	53	46/69
Glov_2063	*mtrA*	9	YP_001228914	*Geobacter uraniireducens* Rf4	66	465/604
Glov_2292	*mtrF*	10	ZP_05313572	*Geobacter* sp. M18	43	473/835
Glov_2294		8	YP_002140171	*Geobacter bemidjiensis* Bem	57	178/245
Glov_2295		9	YP_002537141	*Geobacter* sp. FRC-32	38	283/575
Glov_2299	*mtrF*	10	YP_525308	*Rhodoferax ferrireducens* T118	52	267/396
Glov_2651		1	YP_383296	*Geobacter metallireducens* GS-15	33	69/143
Glov_2758		3	YP_383303	*Geobacter metallireducens* GS-15	52	57/94
Glov_2825		9	YP_002138815	*Geobacter bemidjiensis* Bem	71	503/607
Glov_2826		2	YP_901075	*Pelobacter propionicus* DSM 2379	57	75/101
Glov_3048	*mtrF*	10	ZP_05310497	*Geobacter* sp. M18	43	406/716
Glov_3459		2	YP_903254	*Pelobacter propionicus* DSM 2379	61	86/111
Glov_3541		4	ZP_05311743	*Geobacter* sp. M18	72	481/577
Glov_3625		4	YP_002140822	*Geobacter bemidjiensis* Bem	80	509/576

*Similarity shown as a fraction of positives to aligned length of query to BlastP match

B.8 Table S4. Inferred *c*-type cytochrome genes in *Pelobacter propionicus.*

Locus	Gene symbol	CxxCH motifs	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Chromosome						
Ppro_0721		1	YP_385520	*Geobacter metallireducens* GS-15	75	144/163
Ppro_0828		1	NP_954314	*Geobacter sulfurreducens* PCA	67	112/137
Ppro_0867		5	YP_383646	*Geobacter metallireducens* GS-15	62	365/470
Ppro_1199	*nrfH*	4	YP_001951286	*Geobacter lovleyi* SZ	66	122/153
Ppro_1200	*nrfA*	5	YP_383264	*Geobacter metallireducens* GS-15	70	387/468
Ppro_1401		2	YP_001953059	*Geobacter lovleyi* SZ	57	75/101
Ppro_1575		18	YP_001231869	*Geobacter uraniireducens* Rf4	44	269/465
Ppro_1577		7	YP_385871	*Geobacter metallireducens* GS-15	54	252/366
Ppro_1588		8	YP_001951104	*Geobacter lovleyi* SZ	71	400/470
Ppro_1603		1	YP_001951979	*Geobacter lovleyi* SZ	60	117/155
Ppro_1916		1	YP_958531	*Marinobacter aquaeolei* VT8	99	157/157
Ppro_2999		7	YP_004199549	*Geobacter* sp. M18	77	477/537
Ppro_3498		6	YP_001952039	*Geobacter lovleyi* SZ	67	310/389
Ppro_3509		3	YP_383303	*Geobacter metallireducens* GS-15	67	67/86
Ppro_3605		2	YP_001953682	*Geobacter lovleyi* SZ	61	86/111
Plasmid pPRO1						
Ppro_3655		1	YP_002538020	*Geobacter* sp. FRC-32	56	106/139
Ppro_3705	*nrfH*	5	YP_001229450	*Geobacter uraniireducens* Rf4	80	417/463
Ppro_3707	*nrfA*	4	YP_001229449	*Geobacter uraniireducens* Rf4	70	127/152
Plasmid pPRO2						
Ppro_3831		1	YP_958531	*Marinobacter aquaeolei* VT8	99	157/157
Ppro_3848		1	YP_958531	*Marinobacter aquaeolei* VT8	99	157/157

B.9 Table S5. Electron-transfer proteins encoded on the *G. lovleyi* strain SZ chromosome.

Locus	Function	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Proteins containing molybdopterin oxidoreductase-domain (pfam00384 - underlined) and associated respiratory-type proteins					
Glov_0203	Nitrate reductase, alpha subunit	YP_383297	*Geobacter metallireducens* GS-15	83	1082/1191
Glov_0204	Nitrate reductase, beta subunit	YP_383298	*Geobacter metallireducens* GS-15	78	417/483
Glov_0205	Nitrate reductase molybdenum cofactor assembly	YP_383299	*Geobacter metallireducens* GS-15	58	121/172
Glov_0206	Respiratory nitrate reductase, gamma subunit	YP_383300	*Geobacter metallireducens* GS-15	71	186/221
Glov_0661	Molybdopterin oxidoreductase	YP_002137483	*Geobacter bemidjiensis* Bem	51	545/823
Glov_0899	Formate dehydrogenase	ZP_01311645	*Desulfuromonas acetoxidans* DSM 684	79	669/753
Glov_0900	Iron-sulfur oxidoreductase	YP_519758	*Desulfitobacterium hafniense* Y51	72	97/111
Glov_0930	Molybdopterin oxidoreductase	YP_519751	*Desulfitobacterium hafniense* Y51	63	559/722
Glov_0933	NrfD polysulfide reductase	ZP_01311643	*Desulfuromonas acetoxidans* DSM 684	58	237/314
Glov_0934	4Fe-4S ferredoxin	ZP_01311642	*Desulfuromonas acetoxidans* DSM 684	72	148/181
Glov_0935	Molybdopterin oxidoreductase	ZP_01311641	*Desulfuromonas acetoxidans* DSM 684	78	671/747
Glov_1055	NapD family	YP_001767985	*Methylobacterium* sp. 4-46	32	42/73
Glov_1056	Periplasmic nitrate reductase	ZP_01105283	*Flavobacteriales bacterium* HTCC2170	53	535/773
Glov_1057	NapG ferredoxin	YP_001500485	*Shewanella pealeana* ATCC 700345	48	136/199
Glov_1058	NapH ferredoxin	YP_732839	*Shewanella* sp. MR-4	43	156/249
Glov_1148	4Fe-4S ferredoxin	ZP_03904963	*Denitrovibrio acetiphilus* DSM 12809	62	182/250
Glov_1149	Molybdopterin oxidoreductase	YP_001229259	*Geobacter uraniireducens* Rf4	80	761/867
Glov_1162	NrfD polysulfide reductase	YP_002248979	*Thermodesulfovibrio yellowstonii* DSM 11347	46	221/327
Glov_1163	4Fe-4S ferredoxin	YP_002248980	*Thermodesulfovibrio yellowstonii* DSM 11347	61	146/181
Glov_1164	Formate dehydrogenase	YP_002248981	*Thermodesulfovibrio yellowstonii* DSM 11347	58	546/727
Glov_1443	Molybdopterin oxidoreductase	ZP_03735695	*Desulfonatronospira thiodismutans* ASO3-1	67	330/410

B.9 Table S5. (continued)

Locus	Function	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Proteins containing molybdopterin oxidoreductase-domain (pfam00384 - underlined) and associated respiratory-type proteins (continued)					
Glov_2652	Nitrate reductase, alpha subunit	YP_383297	*Geobacter metallireducens* GS-15	76	1044/ 1193
Glov_2653	Nitrate reductase, beta subunit	YP_001950455	*Geobacter lovleyi* SZ	71	404/481
Glov_2654	Nitrate reductase molybdenum cofactor assembly	YP_383299	*Geobacter metallireducens* GS-15	52	114/173
Glov_2655	Respiratory nitrate reductase, gamma subunit	YP_383300	*Geobacter metallireducens* GS-15	59	169/227
Glov_3550	Molybdopterin oxidoreductase	YP_001230059	*Geobacter uraniireducens* Rf4	60	491/661
Proteins containing flavoprotein domain (pfam00890 - underlined) and associated respiratory-type proteins					
Glov_0943	Flavocytochrome c	YP_002460220	*Desulfitobacterium hafniense* DCB-2	53	358/517
Glov_0945	Succinate dehydrogenase	ZP_01313445	*Desulfuromonas acetoxidans* DSM 684	68	413/509
Glov_0946	Fumarate reductase flavoprotein subunit	ZP_01313446	*Desulfuromonas acetoxidans* DSM 684	50	73/111
Glov_1003	Fumarate reductase/succinate dehydrogenase	YP_002554243	*Diaphorobacter* sp. TPSY	31	238/531
Glov_1298	Fumarate reductase/succinate dehydrogenase	YP_002457273	*Desulfitobacterium hafniense* DCB-2	31	274/583
Glov_2212	4Fe-4S ferredoxin	YP_902144	*Pelobacter propionicus* DSM 2379	84	228/251
Glov_2213	Succinate dehydrogenase flavoprotein	YP_001231990	*Geobacter uraniireducens* Rf4	90	608/638
Glov_2214	Succinate dehydrogenase, cytochrome b	YP_902146	*Pelobacter propionicus* DSM 2379	60	171/222
Glov_3540	C4-dicarboxylate antiporter (Dcu)	ZP_03025442	*Geobacter* sp. M21	88	414/439
Glov_3541	Flavocytochrome c	YP_002140385	*Geobacter bemidjiensis* Bem	72	481/583
Glov_3625	Flavocytochrome c	YP_002140822	*Geobacter bemidjiensis* Bem	79	518/596

B.9 Table S5. (continued)

Locus	Function	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Components to pyruvate flavodoxin/ferredoxin oxidoreductase (PFOR) complexes					
Glov_0368	Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	NP_951159	*Geobacter sulfurreducens* PCA	80	1067/ 1195
Glov_1075	Thiamine pyrophosphate protein domain protein	ZP_03856815	*Thermobaculum terrenum* ATCC BAA-798	63	269/332
Glov_1076	Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	YP_001432409	*Roseiflexus castenholzii* DSM 13941	63	460/592
Glov_1627	Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	YP_001230951	*Geobacter uraniireducens* Rf4	80	341/377
Glov_1628	Thiamine pyrophosphate protein domain protein	YP_384322	*Geobacter metallireducens* GS-15	91	264/272
Glov_1629	Pyruvate ferredoxin/flavodoxin oxidoreductase	NP_952521	*Geobacter sulfurreducens* PCA	80	159/177
Glov_2307	Pyruvate ferredoxin/flavodoxin oxidoreductase	YP_384270	*Geobacter metallireducens* GS-15	72	150/175
Glov_2308	Thiamine pyrophosphate protein domain protein TPP-binding	YP_001230994	*Geobacter uraniireducens* Rf4	80	222/247
Glov_2309	Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	YP_002538449	*Geobacter* sp. FRC-32	76	302/347
Glov_3719 on pSZ77	Thiamine pyrophosphate protein domain protein	YP_158793	*Aromatoleum aromaticum* EbN1	51	232/330
Glov_3720 on pSZ77	Pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	YP_290730	*Thermobifida fusca* YX	54	432/608

B.9 Table S5. (continued)

Locus	Function	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Proteins containing nickel-dependent hydrogenases (pfam00374 - underlined) and associated respiratory-type proteins					
Glov_0144	Hydrogenase small subunit HydA	YP_001230706	*Geobacter uraniireducens* Rf4	79	328/370
Glov_0145	4Fe-4S ferredoxin	YP_001230707	*Geobacter uraniireducens* Rf4	76	267/308
Glov_0146	Polysulphide reductase NrfD	YP_002139935	*Geobacter bemidjiensis* Bem	73	356/412
Glov_0147	Nickel-dependent hydrogenase	NP_951842	*Geobacter sulfurreducens* PCA	78	493/557
Glov_1552	NADH ubiquinone oxidoreductase 20 kDa sub	ZP_03025588	*Geobacter* sp. M21	73	272/308
Glov_1553	Nickel-dependent hydrogenase	YP_002140672	*Geobacter bemidjiensis* Bem	69	398/472
Glov_1554	4Fe-4S ferredoxin	YP_002140671	*Geobacter bemidjiensis* Bem	79	74/82
Glov_2054	Hydrogenase cytochrome b subunit	YP_001229653	*Geobacter uraniireducens* Rf4	57	153/210
Glov_2055	Nickel-dependent hydrogenase	YP_001229331	*Geobacter uraniireducens* Rf4	82	514/565
Glov_2791	Nickel-dependent hydrogenase	YP_846829	*Syntrophobacter fumaroxidans* MPOB	61	370/478
Glov_2792	NADH ubiquinone oxidoreductase, 20 kDa	YP_846828	*Syntrophobacter fumaroxidans* MPOB	57	132/178
Glov_2793	Ferredoxin	YP_384076	*Geobacter metallireducens* GS-15	62	182/231
Glov_2794	NADH dehydrogenase	YP_846826	*Syntrophobacter fumaroxidans* MPOB	65	440/545
Glov_2795	NADH dehydrogenase, 24 kDa	YP_384074	*Geobacter metallireducens* GS-15	61	117/155

B.10 Table S6. Inferred reactive-oxygen-species-responsive or oxygen-reducing genes in *G. lovleyi* strain SZ.

Locus	Function	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Glov_0804	Rubrerythrin	NP_951443	*Geobacter sulfurreducens* PCA	68	128/154
Glov_0828	Rubrerythrin	NP_954329	*Geobacter sulfurreducens* PCA	68	135/165
Glov_1052	Rubrerythrin	YP_386180	*Geobacter metallireducens* GS-15	52	110/149
Glov_1208	Cytochrome bd ubiquinol oxidase subunit I	YP_384884	*Geobacter metallireducens* GS-15	70	371/447
Glov_1209	Cytochrome d ubiquinol oxidase subunit II	NP_952692	*Geobacter sulfurreducens* PCA	74	284/340
Glov_1794	Rubrerythrin	YP_001232869	*Geobacter uraniireducens* Rf4	77	149/168
Glov_1795	Catalase/peroxidase HPI	YP_356594	*Pelobacter carbinolicus* DSM 2380	82	656/735
Glov_1797	Rubrerythrin	YP_901632	*Pelobacter propionicus* DSM 2379	83	168/189
Glov_3343	Ferredoxin	YP_386108	*Geobacter metallireducens* GS-15	67	52/62
Glov_3344	Rubredoxin	ZP_03014104	*Bacteroides intestinalis* DSM 17393	68	41/50
Glov_0804	Rubrerythrin	NP_951443	*Geobacter sulfurreducens* PCA	68	128/154

B.11 Table S7. pSZ77 genes related to plasmid replication, maintenance, and recombination.

pSZ77 locus	Gene symbol	Annotation	Best match to, % identity (similarity)		
			SZ chromosome	*Geobacter* genomes	*Pelobacter* genomes
Glov_3681	*repA*	Plasmid replication protein	None	*G. metallireducens* 14 kbp plasmid 33% (153/303)	*P. propionicus* 31 kbp plasmid 35% (166/293)
Glov_3684*	*parA*	Partitioning ATPase	None	None	None
Glov_3687	*hipB*	Transcriptional regulator, XRE family	None	*Geobacter* sp. M21 chromosome 64% (77/78)	*P. propionicus* 202 kbp plasmid 73% (66/78)
Glov_3688	*hipA*	HipA domain protein	None	*Geobacter* sp. FRC-32 chromosome 28% (159/346)	*P. propionicus* 202 kbp plasmid 75% (344/399)
Glov_3689	*hipB?*	Transcriptional regulator, XRE family	Glov_0180 31% (42/83)	*G. uraniireducens* chromosome 53% (77/112)	*P. propionicus* chromosome 28% (52/100)
Glov_3690	*hipA*	HipA domain protein	Glov_0325 30% (173/406)	*Geobacter* sp. M18 chromosome 65% (331/412)	*P. propionicus* chromosome 29% (168/383)
Glov_3698	*Tn3*	Transposase Tn3	None	None	*P. propionicus* 202 kbp plasmid 46% (640/977)
Glov_3709	-	Integrase	Glov_0344 100%	None	*P. carbinolicus* chromosome 82% (292/324)
Glov_3713	-	Transposase	Glov_2893 100%	None	*P. carbinolicus* chromosome 46% (53/92)
Glov_3714	-	Integrase	Glov_2892 100%	None	None

* The putative pSZ77 ParA protein shares 42% aa identity (61% similarity) with its most similar homolog, a putative ParA on plasmid pMRAD08 from *Methylobacterium radiotolerans* JCM 2831

B.12 Table S8. Inferred cobalamin-dependent genes on the *G. lovleyi* strain SZ chromosome.

Locus	Function	RefSeq ID of top BlastP match	Genome of top BlastP match	% Ident.	Simil.
Glov_2133	Methionine synthase	ZP_08643025	*Brevibacillus laterosporus*	60%	872/ 1161
Glov_2318	Ribonucleotide-diphosphate reductase	YP_902161	*Pelobacter propionicus* DSM 2379	80%	660/735
Glov_2780	PCE reductive dehalogenase	YP_519072	*Desulfitobacterium hafniense* strain Y51 (strain TCE1)	36% (34%)	313/552 (308/550)
Glov_2782	PCE reductive dehalogenase	YP_519072	*Desulfitobacterium hafniense* Y51 (strain TCE1)	35% (34%)	311/552 (308/550)
Glov_3260	Methylmalonyl-CoA mutase	YP_003504649	*Denitrovibrio acetiphilus* DSM 12809	79%	642/712
Glov_0797	Radical SAM domain protein	YP_003657093	*Arcobacter nitrofigilis* DSM 7299	58%	364/484
Glov_1384	Transcriptional regulator, MerR family	YP_003169177	*Candidatus Accumulibacter phosphatis*	41%	165/275
Glov_2185	Radical SAM domain protein	YP_384661	*Geobacter metallireducens* GS-15	74%	397/473
Glov_2195	Radical SAM domain protein	NP_951299	*Geobacter sulfurreducens* PCA	64%	361/459
Glov_2203	Cobalamin B12-binding domain protein	NP_951308	*Geobacter sulfurreducens* PCA	76%	411/471
Glov_2204	Radical SAM domain protein	YP_384656	*Geobacter metallireducens* GS-15	74%	360/419
Glov_2205	Radical SAM domain protein	NP_951302	*Geobacter sulfurreducens* PCA	74%	373/423
Glov_2992	Radical SAM domain protein	YP_383075	*Geobacter metallireducens* GS-15	50%	402/620
Glov_3057	Radical SAM domain protein	YP_902300	*Pelobacter propionicus* DSM 2379	48%	335/516
Glov_3090	Hopanoid biosynthesis associated radical SAM protein	YP_902481	*Pelobacter propionicus* DSM 2379	87%	445/472
Glov_3359	Radical SAM domain protein	YP_356553	*Pelobacter carbinolicus* DSM 2380	32%	228/450
Glov_3363	Radical SAM domain protein	YP_004200859	*Geobacter* sp. M18	51%	538/824
Glov_3401	Radical SAM domain protein	YP_004012404	*Rhodomicrobium vannielii* ATCC 17100	28%	179/387

B.13 Table S9. Plasmid pSZ77 CAI. Codon usage of pSZ77 genes in comparison to strain SZ chromosomal genes using the codon adaptation index (CAI). Out of 79 total predicted plasmid genes, 42 have normalized CAI > 1.00, indicating that over half of pSZ77 genes have resided in the strain SZ genome for sufficient time to ameliorate their codon usage to the chromosome. The predicted replication/maintenance genes *repA* and *parA* have normalized CAI < 1.00 (red font).

pSZ77 locus	Inferred/Annotated Function	Length (bp)	Normalized CAI
Glov_3646	Precorrin-6y C5,15-methyl-transferase, CbiE	618	1.05
Glov_3647	Precorrin-6x reductase, CbiJ/CobK	786	1.07
Glov_3648	Precorrin-3B C17-methyltransferase, CbiH	669	1.15
Glov_3649	Cobalamin biosynthesis protein, CbiG	780	1.02
Glov_3650	Precorrin-4 C11-methyltransferase, CbiF	759	1.05
Glov_3651	Cobalamin biosynthesis protein, CbiD	1077	1.09
Glov_3652	Precorrin-2 C20-methyltransferase, CbiL/CobI	738	1.14
Glov_3653	Anaerobic cobalt chelatase, CbiK	786	1.07
Glov_3654	Precorrin-8X methylmutase, CbiC/CobH	687	1.08
Glov_3655	Cobyrinic acid a,c-diamide synthase, CbiA/CobB	1443	1.11
Glov_3656	Fused siroheme synthase (SirA); Uroporphyrin-III C-methyltransferase (CysG)	1413	1.06
Glov_3657	Cobalt ABC transporter, ATPase subunit	861	1.04
Glov_3658	Cobalt ABC transporter, inner membrane subunit	729	0.94
Glov_3659	Cobalt transport protein, CbiN	333	0.93
Glov_3660	Cobalt transport protein, CbiM	678	1.00
Glov_3661	Hypothetical protein	228	0.87
Glov_3663	RND family efflux transporter	1032	0.85
Glov_3664	S3/IS911 family transposase	288	0.96
Glov_3665	Integrase/Transposase	918	1.04
Glov_3666	Transposase IS4	1329	1.09
Glov_3667	TonB-dependent receptor	1875	1.04
Glov_3668	TonB family protein	735	0.95
Glov_3669	Biopolymer transport protein ExbD/TolR	384	0.84
Glov_3670	TonB system energizer ExbB	435	0.88
Glov_3671	Aerobic cobaltochelatase	4101	1.15
Glov_3672	Uncharacterized domain protein, DUF2149	339	1.15
Glov_3673	MotA/TolQ/ExbB proton channel	597	1.13
Glov_3674	Uncharacterized domain protein, DUF2162	702	1.15
Glov_3675	TonB-dependent receptor	2214	1.10
Glov_3676	O-methyltransferase family protein	1035	1.11
Glov_3677	Hydroxymethyl-pyrimidine synthase, thiC	1311	1.26

B.13 Table S9. (Continued)

pSZ77 locus	Inferred/Annotated Function	Length (bp)	Normalized CAI
Glov_3678	Nicotinate-nucleotide—DMBA phosphoribosyltransferase	1053	1.22
Glov_3679	Fused L-threonine decarboxylase; cobyric acid synthase	2592	1.17
Glov_3680	Histone-like protein	294	0.78
Glov_3681	Replication initiation protein, RepA	1113	0.89
Glov_3682	Hypothetical, no NCBI database hits	195	0.88
Glov_3683	Hypothetical, no NCBI database hits	423	0.91
Glov_3684	Plasmid partitioning ATPase, ParA	651	0.98
Glov_3685	Histone family protein	282	0.79
Glov_3686	Conserved hypothetical protein	2757	1.04
Glov_3687	XRE family transcriptional regulator	240	0.99
Glov_3688	HipA N-terminal domain protein	1212	0.98
Glov_3689	XRE transcriptional regulator	354	1.01
Glov_3690	HipA domain protein	1245	0.98
Glov_3691	WGR domain protein	255	1.08
Glov_3692	Hypothetical, no NCBI database hits	705	0.95
Glov_3693	Hypothetical	510	0.90
Glov_3694	Hypothetical, no NCBI database hits	225	0.87
Glov_3695	Hypothetical	936	0.92
Glov_3696	SAM domain integrase	1014	1.03
Glov_3697	Hypothetical, No NCBI database hits	231	1.01
Glov_3698	Transposase Tn3	2925	1.00
Glov_3700	NADPH-dependent FMN reductase	504	0.94
Glov_3701	Alkylhydroperoxidase	456	1.08
Glov_3702	Hypothetical, No NCBI database hits	276	0.86
Glov_3703	Hypothetical	408	0.87
Glov_3704	TetR family transcriptional regulator	573	1.04
Glov_3705	Hypothetical, No NCBI database hits	336	0.91
Glov_3706	MTH865-like protein	237	1.01
Glov_3707	NADPH-dependent FMN reductase	627	0.95
Glov_3708	Tautomerase	399	1.06
Glov_3709	Integrase	978	1.05
Glov_3710	Hypothetical	588	0.92
Glov_3711	DSBA oxidoreductase	588	1.00
Glov_3712	Short-chain dehydrogenase/reductase	735	0.96
Glov_3713	IS3/IS911 transposase	309	0.90
Glov_3714	Integrase	861	0.92
Glov_3715	IstB domain protein	762	1.14
Glov_3716	Integrase	1572	1.13
Glov_3717	LysR family transcriptional regulator	867	1.00
Glov_3718	Cob(I)alamin adenosyl-transferase, cobA/btuR	663	1.00
Glov_3719	2-Oxoglutarate-ferredoxin-oxidoreductase, beta subunit	1002	1.06

B.13 Table S9. (Continued)

pSZ77 locus	Inferred/Annotated Function	Length (bp)	Normalized CAI
Glov_3720	2-oxoglutarate-ferredoxin-oxidoreductase, alpha subunit	1854	1.08
Glov_3721	Vitamin B12-independent methionine synthase	1098	0.97
Glov_3722	Hypothetical	291	0.90
Glov_3723	Ribosomal small subunit-dependent GTPase A	1002	1.01
Glov_3724	NADPH-dependent FMN reductase	642	0.92
Glov_3725	Crp/FNR family transcriptional regulator	714	1.03
Glov_3726	Haloacid dehalogenase, type II **	687	1.00

* Normalized to codon usage over the entire strain SZ chromosome

** The function of the pSZ77-encoded haloacid dehalogenase is unknown but the protein shares 31% identity (51% similarity) with the characterized 2,2-dichloropropanoic acid dehalogenase from *Pseudomonas putida* strain PP3 (Slater et al., 1985).

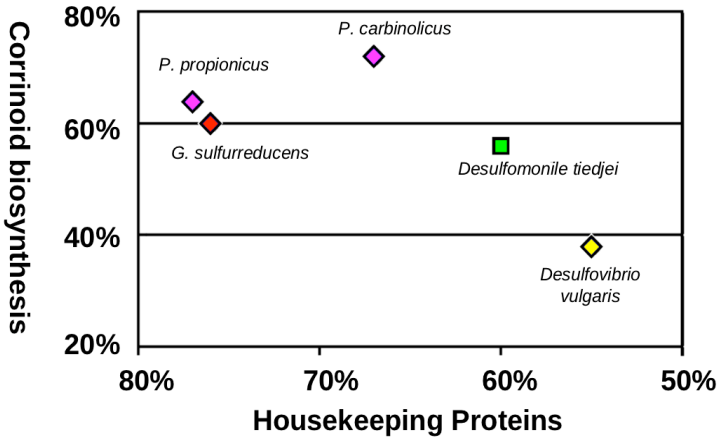
B.14 Table S10. DNA sequence differences between the pSZ77 and the KB-1 plasmid assemblies.

pSZ77 locus	Function	Deletions from KB-1 plasmid
Glov_3646	Precorrin-6y methyltransferase	1 bp deletion 195 bp after start disrupts reading frame on KB-1
Glov_3654	Precorrin-8X methylmutase	136 bp deletion 361 bp from start codon
Glov_3676 to Glov_3675	Intergenic region	4 bp deletion 26 bp upstream of Glov_3675 start 1 bp deletion 10 bp upstream of Glov_3675 start
Glov_3690	HipA	1 bp deletion 717 bp after Glov_3690 start and 81 bp deletion 726 bp after Glov_3690 start, both disrupt reading frame on KB-1

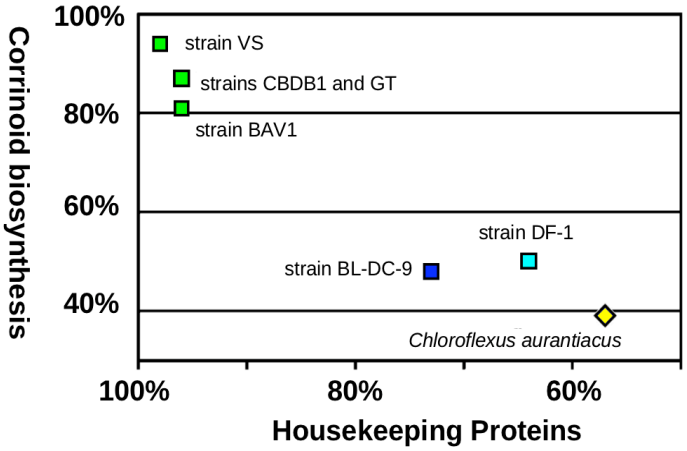
pSZ77 locus	Function	Insertions on KB-1 plasmid not found on pSZ77
Glov_3700	NADPH-dependent FMN reductase	119 bp insertion at 5' end of region corresponding to Glov_3699 pseudogene
Glov_3701	Alkylhydroperoxidase	A 123 bp insertion 80 bp after start of Glov_3701 disrupts reading frame on KB-1
Glov_3712 to Glov_3713 intergenic	Transposase IS3/IS911	A 136 bp insertion 61 bp upstream of Glov_3713

pSZ77 locus	Function	Single Nucleotide Polymorphisms (SNPs)
Glov_3702	Hypothetical, no prokaryotic BlastP hits	T → G changes stop codon to glutamic acid on KB-1 plasmid
Glov_3709	Integrase	Synonymous G → A 360 bp after start codon

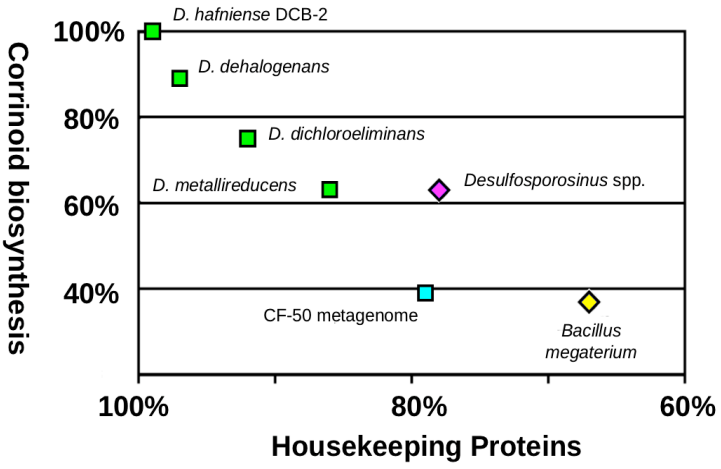
A


Corrinoid biosynthesis
80%
60%
40%
20%
P. carbinolicus
P. propionicus
G. sulfurreducens
Desulfomonile tiedjei
Desulfovibrio vulgaris
80%
70%
60%
50%
Housekeeping Proteins

B


Corrinoid biosynthesis
100%
80%
60%
40%
strain VS
strains CBDB1 and GT
strain BAV1
strain DF-1
strain BL-DC-9
Chloroflexus aurantiacus
100%
80%
60%
Housekeeping Proteins

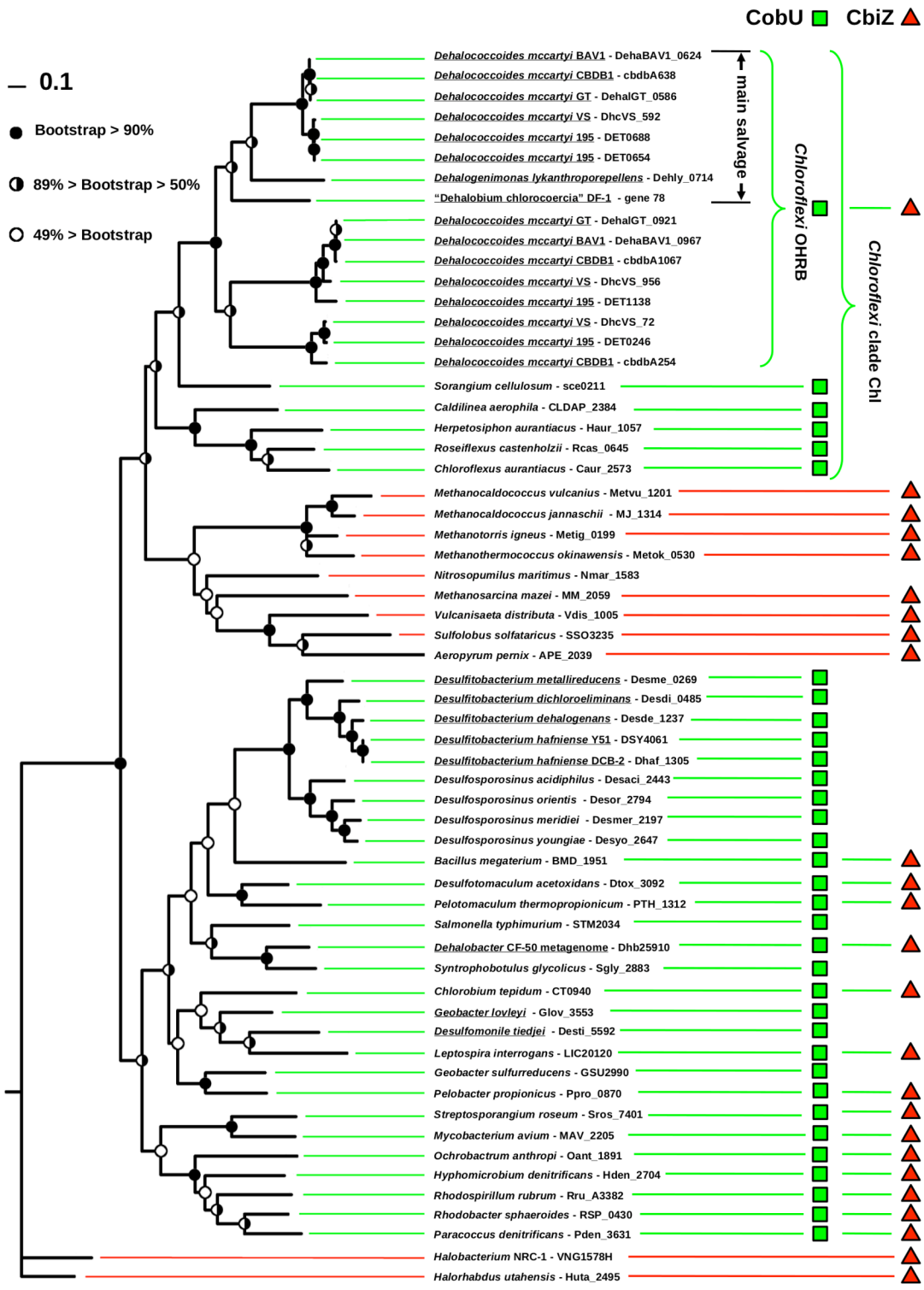
C


Corrinoid biosynthesis
100%
80%
60%
40%
D. hafniense DCB-2
D. dehalogenans
D. dichloroeliminans
D. metallireducens
Desulfosporosinus spp.
CF-50 metagenome
Bacillus megaterium
100%
80%
60%
Housekeeping Proteins

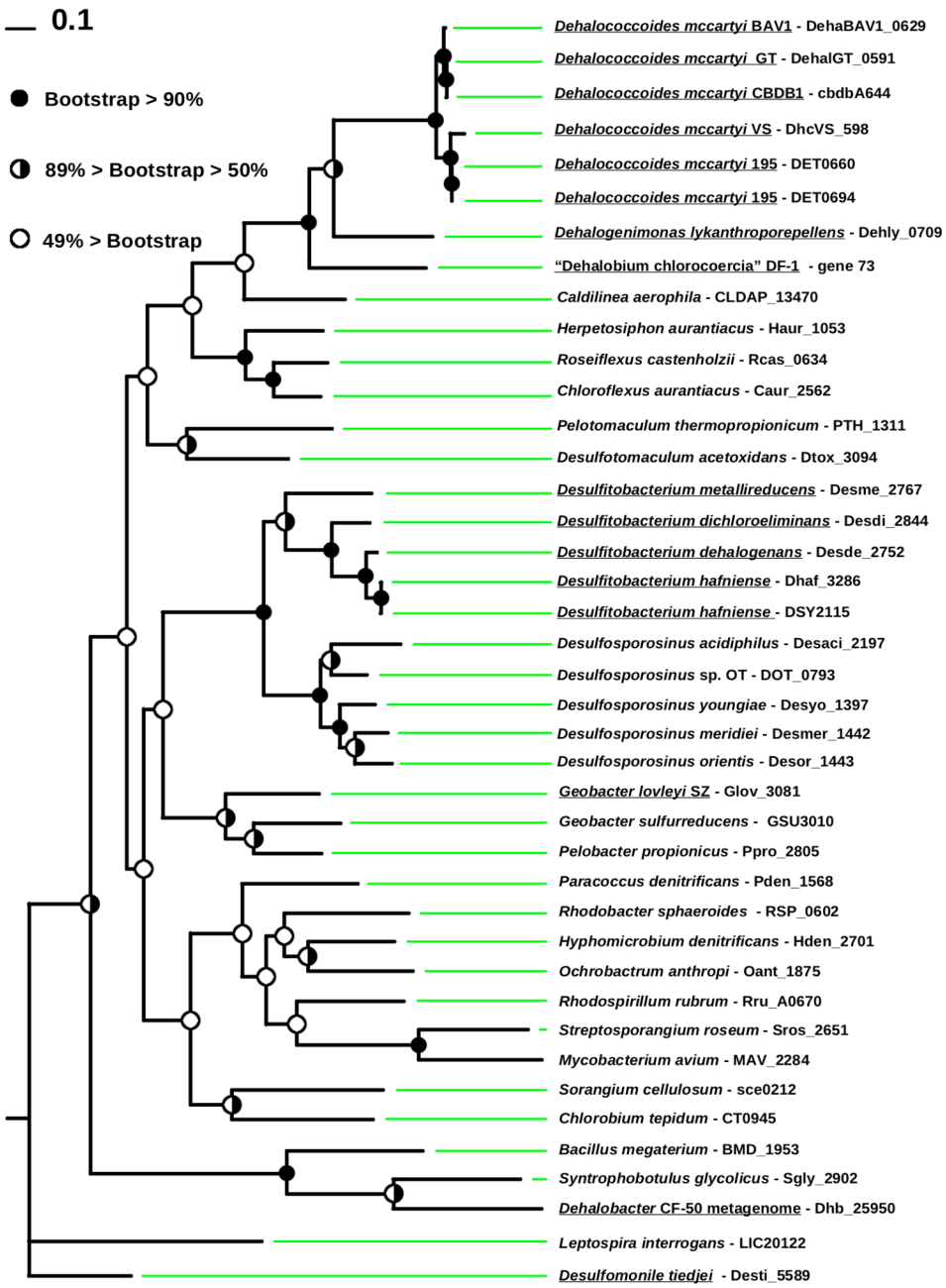
C.1 Figure S1. Corrinoid biosynthesis protein identities, A. Average amino acid identities between *Geobacter lovleyi* strain SZ, *Desulfomonile tiedjei* (green square), and non-organohalide-respiring *δ-Proteobacteria*, *Pelobacter* spp., *Geobacter sulfurreducens*, and *Desulfovibrio vulgaris* (denoted by blue, red, and yellow diamonds, respectively). Percentages along the x-axis represent average identities between housekeeping proteins, DnaA, DnaK, FusA, GltB, GroEL, GyrA, GyrB, RpoB, and TuEF. Percentages along the y-axis represent average identities between corrinoid biosynthesis and transport proteins.

B. Average amino acid identities between model *Chloroflexi* OHRB strain, *Dehalococcoides mccartyi* strain 195 and other members of the *Dehalococcoides* genus (green squares), *Dehalogenimonas* (blue square), "Dehalobium chlorocoercia" (cyan square), and the non-organohalide-respiring *Chloroflexus aurantiacus* (Yellow diamond).

C. Average amino acid identities between *Desulfitobacterium hafniense* strain Y51, other members of the *Desulfitobacterium* genus (green squares), the *Dehalobacter* CF-50 metagenome (cyan square), and non-organohalide-respiring *Firmicutes* (diamonds).


CobU
CbiZ
0.1
Bootstrap > 90%
89% > Bootstrap > 50%
49% > Bootstrap
main salvage
Chloroflexi OHRB
Chloroflexi clade Chl
Dehalococcoides mccartyi BAV1 - DehaBAV1_0624
Dehalococcoides mccartyi CBDB1 - cbdbA638
Dehalococcoides mccartyi GT - DehalGT_0586
Dehalococcoides mccartyi VS - DhcVS_592
Dehalococcoides mccartyi 195 - DET0688
Dehalococcoides mccartyi 195 - DET0654
Dehalogenimonas lykanthroporepellens - Dehly_0714
"Dehalobium chlorocoercia" DF-1 - gene 78
Dehalococcoides mccartyi GT - DehalGT_0921
Dehalococcoides mccartyi BAV1 - DehaBAV1_0967
Dehalococcoides mccartyi CBDB1 - cbdbA1067
Dehalococcoides mccartyi VS - DhcVS_956
Dehalococcoides mccartyi 195 - DET1138
Dehalococcoides mccartyi VS - DhcVS_72
Dehalococcoides mccartyi 195 - DET0246
Dehalococcoides mccartyi CBDB1 - cbdbA254
Sorangium cellulosum - sce0211
Caldilinea aerophila - CLDAP_2384
Herpetosiphon aurantiacus - Haur_1057
Roseiflexus castenholzii - Rcas_0645
Chloroflexus aurantiacus - Caur_2573
Methanocaldococcus vulcanius - Metvu_1201
Methanocaldococcus jannaschii - MJ_1314
Methanotorris igneus - Metig_0199
Methanothermococcus okinawensis - Metok_0530
Nitrosopumilus maritimus - Nmar_1583
Methanosarcina mazei - MM_2059
Vulcanisaeta distributa - Vdis_1005
Sulfolobus solfataricus - SSO3235
Aeropyrum pernix - APE_2039
Desulfitobacterium metallireducens - Desme_0269
Desulfitobacterium dichloroeliminans - Desdi_0485
Desulfitobacterium dehalogenans - Desde_1237
Desulfitobacterium hafniense Y51 - DSY4061
Desulfitobacterium hafniense DCB-2 - Dhaf_1305
Desulfosporosinus acidiphilus - Desaci_2443
Desulfosporosinus orientis - Desor_2794
Desulfosporosinus meridiei - Desmer_2197
Desulfosporosinus youngiae - Desyo_2647
Bacillus megaterium - BMD_1951
Desulfotomaculum acetoxidans - Dtox_3092
Pelotomaculum thermopropionicum - PTH_1312
Salmonella typhimurium - STM2034
Dehalobacter CF-50 metagenome - Dhb25910
Syntrophobotulus glycolicus - Sgly_2883
Chlorobium tepidum - CT0940
Geobacter lovleyi - Glov_3553
Desulfomonile tiedjei - Desti_5592
Leptospira interrogans - LIC20120
Geobacter sulfurreducens - GSU2990
Pelobacter propionicus - Ppro_0870
Streptosporangium roseum - Sros_7401
Mycobacterium avium - MAV_2205
Ochrobactrum anthropi - Oant_1891
Hyphomicrobium denitrificans - Hden_2704
Rhodospirillum rubrum - Rru_A3382
Rhodobacter sphaeroides - RSP_0430
Paracoccus denitrificans - Pden_3631
Halobacterium NRC-1 - VNG1578H
Halorhabdus utahensis - Huta_2495

C.2 Figure S2. Bootstrapped phylogeny of 60 adenosylcobinamide phosphate synthase (CbiB) proteins. Red highlighted leaves denote Archaea and green-highlighted leaves denote bacteria. Green squares denote the presence of bifunctional cobinamide biosynthesis/salvage proteins (CobU) encoded on the genome with CbiB. Similarly, red triangles denote the presence of cobinamide amidohydrolase (CbiZ) on genomes with CbiB. The scale bar indicates change in 10% of aligned amino acid sites. Closed circles indicate at least 90% bootstrap support (out of 100), half-closed circles indicate 50% to 89% bootstrap support, and open circles indicate less than 50% bootstrap support. OHRB genome/strain names are underlined. Note that CbiZ of both OHRB and non-organohalide respirers of *Chloroflexi* cluster with archaeal CbiZ with > 50% support.



C.3 Figure S3. Bootstapped phylogeny of 41 bifunctional cobinamide biosynthesis/salvage protein CobU. CobU = cobinamide kinase/cobinamide phosphate guanylyltransferase, EC:2.7.1.156; EC: 2.7.7.62. The scale bar indicates change in 10% of aligned amino acid sites. Closed circles indicate at least 90% bootstrap support (out of 100), half-closed circles indicate 50% to 89% bootstrap support, and open circles indicate less than 50% bootstrap support. OHRB genome/strain names are underlined.

C.4 Table S1: **OHRB corrinoid biosynthesis and transport gene repertoire.** OHRB with experimental confirmation of *de novo* corrinoid biosynthesis or corrinoid salvage are underlined. The uroporphyrinogen III biosynthesis genes, *hemA-B-C-D*, are not included.

	cysG	*cbiK*	*-L*	*-H*	*-G*	*-F*	*-D*	*-J*	*-E*	*-T*	*-C*	*-A*	*btuR cobA*	*cbiP*	*cobD*	*cbiZ*	*cbiB*	*cobT*	*cobU*	*cobS*	*cobC*	*cbiM-N-Q-O*	*btuB TonB*	*-F*	*-C*	*-D*
***Geobacter lovleyi* strain SZ**	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■
***Anaeromyxobacter dehalogenans* 2CP-C**																							■	■	■	■
***Anaeromyxobacter dehalogenans* 2CP-1**																							■	■	■	■
***Anaeromyxobacter* sp. K**																							■	■	■	■
***Desulfitobacterium hafniense* Y51**	■		■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■			■	■
***Desulfitobacterium hafniense* DCB-2**	■		■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■			■	■
***Dehalobacter* metagenome CF-50**	■		■	■	■	■	■	■	■	■	■	■	■	■	■		■	■	■	■	■	■		■	■	■
***Dehalococcoides ethenogenes* 195**									■				■	■	■	■	■	■	■	■	■			■	■	■
***Dehalococcoides* sp. CBDB1**									■				■	■	■	■	■	■	■	■	■			■	■	■
***Dehalococcoides* sp. BAV1**													■	■	■	■	■	■	■	■	■			■	■	■
***Dehalococcoides* sp. VS (mixed?)**									■				■	■	■	■	■	■	■	■	■			■	■	■
***Dehalococcoides* sp. GT**													■	■	■	■	■	■	■	■	■			■	■	■
***Dehalogenimonas* sp. BL-DC-9**									■				■	■	■	■	■	■	■	■	■			■	■	■
***Dehalobium* sp. DF-1**		■	■	■	■	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■			■	■	■

C.5 Table S2. Corrinoid biosynthesis and salvage genomic similarities. Identities between model OHRB *Geobacter lovleyi* strain SZ versus *Desulfitobacterium hafniense* strain Y51 *Dehalococcoides mccartyi* strain 195, and a non-OHRB, *Geobacter sulfurreducens.*

Geobacter lovleyi gene locus	Gene symbol	Functional annotation	*Geobacter sulfurreducens*	Identity (Similarity)	*Desulfitobacterium hafniense*	Identity (Similarity)	*Dehalococcoides mccartyi* strain 195	Identity (Similarity)
Glov_0503	*hemA*	Glutamyl-tRNA reductase	hemA	73% (88%)	hemA	43% (66%)	no match	no match
Glov_0688	*hemB*	Delta-aminolevulinic acid dehydratase	hemB	84% (92%)	DSY2222	61% (76%)	no match	no match
Glov_0502	*hemC*	Porphobilinogen deaminase	hemC	80% (86%)	hemC	48% (65%)	no match	no match
Glov_0501	*hemD*	Uroporphyrinogen-III synthase	GSU3286	49% (63%)	DSY2224	37% (54%)	no match	no match
Glov_3656	*cysG*	Uroporphyrinogen III C-methyltransferase	GSU3286	47% (61%)	DSY2224	45% (59%)	no match	no match
Glov_3653	*cbiK*	Anaerobic cobalt chelatase	no match	no match	no match	no match	no match	no match
Glov_3652	*cbiL*	Precorrin-2 C20-methyltransferase	cobI	37% (53%)	cobI	38% (52%)	no match	no match
Glov_3648	*cbiH*	Precorrin-3B C17-methyltransferase	cobJQ	55% (68%)	cobJ	48% (68%)	no match	no match
Glov_3650	*cbiF*	Precorrin-4 C11-methyltransferase	cobM	52% (68%)	cobM	55% (68%)	no match	no match
Glov_3649	*cbiG*	Cobalamin biosynthesis CbiG protein	GSU2993	41% (60%)	cbiG	38% (59%)	no match	no match
Glov_3651	*cbiD*	Cobalamin biosynthesis protein CbiD	cbiD	42% (55%)	cbiD	43% (62%)	no match	no match
Glov_3647	*cbiJ*	Precorrin-6x reductase	no match	no match	cobK	34% (50%)	no match	no match
Glov_3646	*cbiE*	Precorrin-6y C5,15-methyltransferase	cobL	50% (65%)	cobL	33% (55%)	no match*	no match*
Glov_3417	*cbiT*	Precorrin decarboxylase	cobL	70% (85%)	cobL	31% (50%)	no match	no match
Glov_3654	*cbiC*	Precorrin-8X methylmutase CbiC/CobH	cobH	48% (62%)	cobH	38% (54%)	no match	no match
Glov_3655	*cbiA*	Cobyrinic acid a,c-diamide synthase	cobB	44% (60%)	cobB	39% (58%)	cobB	40% (57%)
Glov_0692	*btuR*	Cob(I)alamin adenosyltransferase	cobO	69% (80%)	cobA	50% (69%)	cobA-3	44% (63%)
Glov_3554	*cbiP*	Cobyric acid synthase (c-terminal)	cobJQ	45% (63%)	cobQ	48% (63%)	cobQ	46% (64%)

C.5 Table S2: (Continued)

Geobacter lovleyi gene locus	Gene symbol	Functional annotation	*Geobacter sulfurreducens*	Identity (Similarity)	*Desulfitobacterium hafniense*	Identity (Similarity)	*Dehalococcoides mccartyi* strain 195	Identity (Similarity)
Glov_3554	*cobD*	L-threonine O-3-phosphate decarboxylase (n-terminal)	GSU2989	38% (57%)	DSY4057	32% (50%)	DET0689	32% (51%)
Glov_3553	*cbiB*	Adenosylcobinamide-phosphate synthase	cobD	56% (66%)	cobD	43% (57%)	cobD-3	35% (50%)
Glov_3081	*cobU*	Adenosylcobinamide kinase (EC 2.7.1.156)	cobU	57% (70%)	cobU	43% (59%)	cobU-2	35% (59%)
Glov_3082	*cobT*	Nicotinate-nucleotide dimethylbenzimidazole	cobT	73% (85%)	cobT	52% (71%)	cobT-2	51% (69%)
Glov_3084	*cobC*	Alpha-ribazole phosphatase	GSU3007	58% (74%)	cobC	37% (58%)	DET0693	32% (52%)
Glov_3083	*cobS*	Cobalamin-5'-phosphate synthase (EC 2.7.8.26)	cobS	52% (66%)	cobS	37% (55%)	cobS-2	36% (52%)
Glov_3557	*btuF*	Periplasmic cobalamin binding protein	no match	no match	DSY2084	27% (47%)	DET0684	30% (52%)
Glov_3558	*btuC*	Fe3+-siderophore transporter, permease	no match	no match	DSY2085	40% (60%)	DET0685	43% (64%)
Glov_3559	*btuD*	Fe3+-siderophore transporter, ATP-binding	GSU3001	34% (56%)	DSY1005	41% (57%)	DET0686	32% (54%)
Glov_3657	*cbiO*	Cobalt ABC transporter, ATPase subunit	GSU3001	38% (57%)	cbiO	44% (64%)	DET0417	32% (56%)
Glov_3658	*cbiQ*	Cobalt ABC transporter, inner membrane subunit CbiQ	GSU3002	28% (49%)	DSY1481	25% (52%)	no match	no match
Glov_3659	*cbiN*	Cobalt transport protein CbiN	no match	no match	DSY1480	58% (63%)	no match	no match
Glov_3660	*cbiM*	Cobalamin biosynthesis protein CbiM	GSU3004	37% (54%)	cbiM	69% (80%)	no match	no match

C.6 Table S3: Corrinoid salvage and transport gene loci on *Chloroflexi* OHRB.

Genome Region	Adjacent corrin-dependent genes	Corrinoid metabolism gene	*D. mccartyi* strain 195	*D. mccartyi* strain CBDB1	*D. mccartyi* strain BAV1	*D. mccartyi* strain GT	*D. mccartyi* strain VS	*Dehalogenimonas* strain BL-DC-9
Genomic islands*	RDases	***cbiZ***	DET0314, DET1165, DET1556	cbdbA87, cbdbA81, cbdbA1097 cbdbA1499, cbdbA1554	DehaBAV1_0102 DehaBAV1_0993	DehalGT_0121 DehalGT_1280	DhcVS_1356	
	Ribonucleotide reductase	***cbiE***	DET0240, DET0296	cbdbA247			DhcVS_78	Dehly_0579
		cbiZ	DET0242	cbdbA249			DhcVS_76	
		cobA	DET0245	cbdbA253			DhcVS_73	Dehly_0580
		cbiB	DET0246	cbdbA254			DhcVS_72	
		cbiZ	DET0249	cbdbA258			DhcVS_69	
		btuF/ cbiE	DET0250	cbdbA259			DhcVS_68	
Core genomic	Unknown	***cbiA***	DET0128	cbdbA149	DehaBAV1_0243	DehalGT_0156	DhcVS_135	Dehly_0474, Dehly_0475, Dehly_0610
	Carbon monoxide dehydrogenase/ acetyl-CoA synthase	***btuF***	DET0650, DET0684	cbdbA633	DehaBAV1_0620	DehalGT_0582	DhcVS_588	Dehly_0718, Dehly_0717
		btuC	DET0651, DET0685	cbdbA635	DehaBAV1_0621	DehalGT_0583	DhcVS_589	Dehly_0716
		btuD	DET0652, DET0686	cbdbA636	DehaBAV1_0622	DehalGT_0584	DhcVS_590	Dehly_0715
		cbiZ	DET0653, DET0687	cbdbA637	DehaBAV1_0623	DehalGT_0585	DhcVS_591	Dehly_0902
		cbiB	DET0654, DET0688	cbdbA638	DehaBAV1_0624	DehalGT_0586	DhcVS_592	Dehly_0714
		cobD	DET0655, DET0689	cbdbA639	DehaBAV1_0625	DehalGT_0587	DhcVS_593	Dehly_0713
		cobT	DET0657, DET0691	cbdbA641	DehaBAV1_0626	DehalGT_0588	DhcVS_595	Dehly_0712
		cobS	DET0658, DET0692	cbdbA642	DehaBAV1_0627	DehalGT_0589	DhcVS_596	Dehly_0711
		cobC	DET0659, DET0693	cbdbA643	DehaBAV1_0628	DehalGT_0590	DhcVS_597	Dehly_0710
		cobU	DET0660, DET0694	cbdbA644	DehaBAV1_0629	DehalGT_0591	DhcVS_598	Dehly_0709
	Unknown	***cbiP***	DET0936	cbdbA890	DehaBAV1_0820	DehalGT_0781	DhcVS_807	Dehly_0845
	Unknown	***cobA***	DET1138	cbdbA1069	DehaBAV1_0968	fragment	DhcVS_957	
	Unknown	***cbiB***	DET1139	cbdbA1067	DehaBAV1_0967	DehalGT_0921	DhcVS_956	
	Unknown	***cobA***	DET1224	cbdbA1141	DehaBAV1_1033	DehalGT_0964	DhcVS_1006	Dehly_1006

C.7 Table S4. Numbers of corrinoid-dependent genes on OHRB genomes. For draft genomes, absence of homologs could not be verified and are indicated as "n/d" (not detected).

Genome	Known RDases	Putative RDases	Ribonucleotide reductase	Acetyl-CoA synthase, delta or gamma subunit	Methylmalonyl -CoA mutase	Methionine synthase	Demethylase-associated proteins	Ethanol-amine lyase, small subunit	Hypothetical corrinoid-binding
Geobacter lovleyi strain SZ	.	2	1	0	1	1	0	0	12
Desulfomonile tiedjei strain DCB-1	.	3	1	1	1	0	14	0	0
Anaeromyxobacter dehalogenans strain 2CP-C	.	2	1	0	1	1	0	0	1
Anaeromyxobacter dehalogenans strain 2CP-1	.	2	1	0	1	1	0	0	2
Anaeromyxobacter dehalogenans strain K	.	2	1	0	1	1	0	0	2
Desulfitobacterium hafniense strain Y51	1 PceA	0	1	1	1	1	15	1	1
Desulfitobacterium hafniense strain DCB-2	1 CprA	6	1	1	1	1	15	1	1
Desulfitobacterium dehalogenans	1 CprA	5	1	1	1	1	10	0	1
Desulfitobacterium dichloroeliminans	1 DcaA	0	1	1	1	1	4	0	n/d
Desulfitobacterium metallireducens	.	n/d	1	1	1	1	n/d	0	1
Dehalobacter CF-50 metagenome	.	17	1	1	n/d	1	n/d	n/d	n/d
Dehalococcoides mccartyi strain 195	1 PceA, 1 TceA	15	1	1	0	0	0	0	0
Dehalococcoides mccartyi strain VS	1 VcrA	35	1	1	0	0	0	0	0
Dehalococcoides mccartyi strain CBDB1	2 CbrA	30	1	1	0	0	0	0	0
Dehalococcoides mccartyi strain BAV1	1 BvcA	10	1	1	0	0	0	0	0
Dehalococcoides mccartyi strain GT	.	20	1	1	0	0	0	0	0
Dehalogenimonas lykanthroporepellens strain BL-DC-9	.	19	1	1	0	0	0	0	0
Candidatus Dehalobium chlorocoercia strain DF-1	.	30	n/d	1	n/d	1	n/d	n/d	n/d